JC17 Rec'd PCT/PTO 0 8 MAY 2001

FORM PTO-1390 (REV 1-98) U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER: 146.1364

U.S. APPLICATION NO. (If known, see 37 CFR 1.5): 09/831426

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/FR99/02738	November 9, 1999	November 10, .1998

TITLE OF INVENTION: HUMAN htFIIIA GENE AND CODED htFIIIA PROTEIN

APPLICANT(S) FOR DO/EO/US: BORDON-PALLIER et al

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a FIRST submission of items concerning a filing under 35 U.S.C. 371.
2. [ ] This is a SECOND or SUBSEQUENT submission of items concerning a filing under 35 U.S.C. 371.
3. [X] This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. [ ] A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. [X] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. [X] is transmitted herewith (required only if not transmitted by the International Bureau).
   b. [ ] has been transmitted by the International Bureau.
   c. [ ] is not required, as the application was filed in the United States Receiving Office (RO/US).
6. [X] A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. [ ] Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. [ ] are transmitted herewith (required only if not transmitted by the International Bureau).
   b. [ ] have been transmitted by the International Bureau.
   c. [ ] have not been made; however, the time limit for making such amendments has NOT expired.
   d. [ ] have not been made and will not be made.
8. [X] A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).
9. [X] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. [ ] A translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11. to 16. below concern document(s) or information included:

11. [X] An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. [ ] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. [X] A FIRST preliminary amendment.
    [ ] A SECOND or SUBSEQUENT preliminary amendment.
14. [ ] A substitute specification.
15. [ ] A change of power of attorney and/or address letter.
16. [X] Other items or information: Letter re: Sequence Listing; Paper Copy of Sequence Listing and Diskette Corresponding thereto; International Preliminary Examination Report; PCT/IB/306; PTO Form 2038

JC08 Rec'd PCT/PTO 0 8 MAY 2001

U.S. APPLICATION NO. (if known, see 37 CFR 1.5) 09/831426	INTERNATIONAL APPLICATION NO. PCT/FR99/02738	ATTORNEY'S DOCKET NUMBER 146.1364

17. [x] The following fees are submitted:

	CALCULATIONS PTO USE ONLY	
BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):		
Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO $1070.00	$1000.00	
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO $930.00		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO $790.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) $720.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) $98.00		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$1000.00	
Surcharge of $130.00 for furnishing the oath or declaration later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total claims	- 20 =		x $22.00	$	
Independent claims	- 3 =		x $82.00	$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ $270.00	$	
TOTAL OF ABOVE CALCULATIONS =				$1000.00	
Reduction of 1/2 for filing by small entity, if applicable. A Small Entity Statement must also be filed (Note 37 CFR 1.9, 1.27, 1.28). +				$	
SUBTOTAL =				$1000.00	
Processing fee of $130.00 for furnishing the English translation later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				$	
TOTAL NATIONAL FEE =				$1000.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). $40.00 per property +				$	
TOTAL FEES ENCLOSED =				$1000.00	
				Amount to be refunded:	$
				charged:	$

a. [x] PTO Form 2038 is enclosed ~~A check in the amount of $ _____ to cover the above fees is enclosed.~~

b. [ ] Please charge my Deposit Account No. _____ in the amount of $ _____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. [x] The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 02-2275 A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Bierman, Muserlian and Lucas
600 Third Avenue
New York, NY 10016

SIGNATURE

Charles A. Muserlian
NAME

19,683
REGISTRATION NUMBER

05-09-01

09/831426 PCT 8 Sea.

JC08 Rec'd PCT/PTO 08 MAY 2001

146.1364

"EXPRESS MAIL" Mailing Label Number EL 783039655 US

Date of Deposit: May 8, 2001

I hereby certify that this correspondence is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" Service under 37 CFR 1.10 on the date indicated above and is addressed to Asst. Commissioner for Patents, Washington, D.C. 20231.

Charles A. Muserlian

146.1364

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	:	PCT Date: 11/9/99
F. BORDON-PALLIER et al	:	
PCT No.: PCT/FR99/02738	:	
Filed: Concurrently Herewith	:	
For: HUMAN...PROTEIN	:	

600 Third Avenue
New York N.Y. 10016

PRELIMINARY AMENDMENT

Asst. Commissioner for Patents
Washington, D.C. 20231

Sir:

Please amend this application as follows:

IN THE SPECIFICATION:

Page 1, before line 1, insert

--This application is a 371 of PCT/FR99/02738 filed November 9,1999.--

IN THE CLAIMS:

Claim 3 (amended) DNA sequence of the htfIIIA gene according to claim 1 containing the nucleotide sequence SEQ ID No: 3.

Claim 4 (amended) DNA sequence of the htfIIIA gene according to claim 1 containing the nucleotide sequence SEQ ID No: 4.

Claim 5 (amended) DNA sequence according to claim 4 having the sequence beginning at nucleotide 176 and finishing at the nucleotide 1270 of SEQ ID No: 3.

Claim 6 (amended) DNA sequence coding for the human transcription factor hTFIIIA according to claim 1 as well as the

DNA sequence which hybridize with it and/or show a significant homology with this sequence or fragments of it and which code for a protein with the same function.

Claim 7 (amended) DNA sequence according to claim 1 comprising modifications introduced by suppression, insertion and/or substitution of at least one nucleotide coding for a protein with the same biological activity as human transcription factor hTFIIIA.

Claim 8 (amended) DNA sequence according to claim 1 as well as similar DNA sequences which have nucleotide sequence homology of at least 50% or at least 60% and preferably at least 70% with the said DNA sequence.

Claim 9 (amended) DNA sequence according to claim 1 as well as similar DNA sequences which code for a protein, the AA sequence of which has a homology of at least 40% and in particular 45% or at least 50%, rather at least 60% and preferably at least 70% with the AA sequence coded by the said DNA sequence.

Claim 10 (amended) Polypeptide having the function of human transcription factor hTFIIIA and with the amino acid sequence SEQ ID No: 2 coded by the DNA sequence according to claim 1 and the analogues of this polypeptide.

Claim 11 (amended) Process for the preparation of the hTFIIIA recombinant protein having the amino acid sequence SEQ ID No: 2 comprising the expression of the DNA sequence according to claim 1 in an appropriate host, then isolation and purification of the said recombinant protein.

Claim 12 (amended) Expression vector containing the DNA sequence according to claim 3.

Cancel claims 15 and 16 and add the following claims:

--17. A method of treating a disease linked to transcription control disorders in warm-blooded animals comprising administering to warm-blooded animals in need thereof an amount of the DNA sequence of claim 1 or the human transcription factor coded by the sequence sufficient to treat said diseases.

18. The method of claim 17 wherein the disease is cancer.--

REMARKS

The amendment is presented to insert reference to the PCT application, to remove multiple dependency from the claims and to present proper method of use claims. A marked up copy of the amended claims is submitted herewith.

Respectfully submitted,
Bierman, Muserlian and Lucas

By: 
Charles A. Muserlian #19,683
Attorney for Applicants
Tel.# (212) 661-8000

CAM:ds
Enclosures

CLAIMS

1) DNA sequence of the htfIIIA gene coding for a protein having the biological function of human transcription factor hTFIIIA.

2) DNA sequence of the htfIIIA gene of the human transcription factor hTFIIIA according to claim 1, coding for the amino acid sequence SEQ ID N°2.

3) DNA sequence of the htfIIIA gene according to claim 1 ~~or 2~~ containing the nucleotide sequence SEQ ID N°3

4) DNA sequence of the htfIIIA gene according to claims 1 ~~to~~ containing the nucleotide sequence SEQ ID N°4.

5) DNA sequence according to claim 4 having the sequence beginning at nucleotide 176 and finishing at the nucleotide 1270 of SEQ ID N°3.

6) DNA sequence coding for the human transcription factor hTFIIIA according to claims 1 ~~to 5~~ as well as the DNA sequences which hybridize with it and/or show a significant homology with this sequence or fragments of it and which code for a protein with the same function.

7) DNA sequence according to claims 1 ~~to 6~~ comprising modifications introduced by suppression, insertion and/or substitution of at least one nucleotide coding for a protein with the same biological activity as human transcription factor hTFIIIA.

8) DNA sequence according to ~~one of~~ claims 1 ~~to 7~~ as well as similar DNA sequences which have nucleotide sequence homology of at least 50 % or at least 60 % and preferably at least 70 % with the said DNA sequence.

9) DNA sequence according to ~~one of~~ claims 1 ~~to 8~~ as well as similar DNA sequences which code for a protein, the AA sequence of which has a homology of at least 40 % and in particular 45 % or at least 50 %, rather at least 60 % and preferably at least 70 % with the AA sequence coded by the said DNA sequence.

10) Polypeptide having the function of human transcription factor hTFIIIA and with the amino acid sequence SEQ ID N°2 coded by the DNA sequence according to ~~one of~~ claims 1 ~~to 9~~ and the analogues of this polypeptide.

11) Process for the preparation of the hTFIIIA recombinant protein having the amino acid sequence SEQ ID N°2 comprising the expression of the DNA sequence according to ~~one of~~ claims 1 ~~to 9~~ in an appropriate host, then isolation and purification of the said recombinant protein.

12) Expression vector containing the DNA sequence according to ~~one of~~ claims 3 ~~to 9~~.

13) Host cell transformed with a vector according to claim 12

14) Plasmid deposited at the CNCM under the number I-2071.

15) Use of the human transcription factor htfIIIA gene or of the human transcription factor coded by this gene according to one of the claims 1 to 10 for the preparation of compositions which can be used for the diagnosis or treatment of diseases linked to a disorder in transcription control.

16) Use according to claim 15 for which the disease concerned is cancer.

Human htFIIIA gene and coded htfIIIA protein

The present invention relates to the gene coding for the human transcription factor hereafter called htFIIIA (or htfC2) gene and the coded htfIIIA protein, as well as the use of this htFIIIA gene and that of the coded htfIIIA protein in the diagnosis and identification of certain diseases related to the transcription mechanism.

Hereafter the gene coding for the transcription factor TFIIIA will be called tfIIIA (or tfC2) and the gene coding for the human transcription factor htfIIIA will be called htfIIIA.

The human htFIIIA gene codes therefore for the corresponding htFIIIA protein.

We will also use the following abbreviations below: AA for amino acids, NA for nucleic acids, bp for base pairs, DNA for deoxyribonucleic acid, cDNA for complementary DNA, RNA for ribonucleic acid, RNase for ribonuclease and C for deoxycytidine.

The term screening which indicates a specific screening technique and the term primer which indicates an oligonecleotide used as primer will also be used.

The tfIIIA gene and the corresponding tfIIIA protein are involved in the regulation of the biological transcription mechanism as indicated below.

Since the tfIIIA protein was purified as transcription factor for the first time in 1980 from Xenopus ovocytes [Segall et al, J. Biol. Chem., 255, 11986-11991 (1980)], work has been carried out in vivo and in vitro within the Xenopus to study the mechanism of transcription control exercised by TFIIIA. It has thus been shown that Xenopus TFIIIA is necessary for the initiation of the transcription of 5S RNA gene [Sakonji et al, Cell 19, 13-25 (1980)] and binds to a internal control region of the 5S RNA gene [Bogenhagen et al, Cell, 19,27-35 (1980)].

The nucleotide sequence of the cDNA of Xenopus TFIIIA and the corresponding amino acid sequence have already been published [Ginberg et al, Cell 39,479-489 (1984)]. It can be

noted that this gene codes for a structure of 9 zinc fingers, a zinc finger corresponding to the repetition of the CYS2 HIS2 (C2H2) moiety. This zinc finger structure is considered an essential domain for a group of proteins which bind themselves to the DNA (DNA binding proteins) [Miller et al, Embo J., 4, 1607-1614 (1985)].

In this way transcription factors in human beings, binding to the DNA which also have this zinc finger structure such as for example XT1 of the Wilms human tumor gene, [Gessier et al, Nature, 343, 774-778 (1990)], the YY1 human transcription repressor [SHI et al, Cell, 67, 377-388 (1991)], the MAZ protein combined with the human MYC gene [Bossone et al, Proc. Natl. Acad. Sci., USA, 89, 7452-7456 (1992)] or also sp1 [Kuwahara et al, J.Biol. Chem., 29, 8627-8631 (1990)] are known.

Studies have been carried out in order to isolate the human htFIIIA gene, but until now none have led to discovery of the true sequence of the htFIIIA gene.

On one hand the studies described in the European Application EP 0704526 (Fujisawa et al), can thus be mentioned and are examined in the article: Arakawa et al (1995), Cytogenet Cell Genet 70, 235-238, which have led to a sequence that we will call Arakawa htfIIIA and on the other hand the studies described in the article: DREW et al (1995), Gene 159, 215-218, which have led to a sequence that we will call DREW htfIIIA. These DREW and ARAKAWA htfIIIA sequences are represented in Figures 4 and 5 respectively below. The documents indicated above therefore each describe a sequence of the htfIIIA gene but these two sequences differ from one another by a few nucleotides and differ from the htfIIIA gene of the present Application as indicated below.

The present invention has made it possible to isolate the gene coding for the human transcription factor hTFIIIA.

The present invention has also made it possible to reveal the nucleic acid sequence of the htfIIIA gene and also the amino acid sequence of the hTFIIIA protein coded by this gene.

Therefore a subject of the present invention is the DNA

sequence of the htfIIIA gene coding for a protein having the biological function of human transcription factor htfIIIA.

A precise subject of the present invention is the DNA sequence of the htfIIIA gene of human transcription factor hTFIIIA as defined above, coding for the amino acid sequence SEQ ID N°2.

Such a SEQ ID n°2 sequence of the present invention therefore comprises 365 amino acids.

A subject of the present invention is also the DNA sequence of the htfIIIA gene as defined above, containing the nucleotide sequence SEQ ID N°3.

A subject of the present invention is the DNA sequence of the htfIIIA gene as defined above, containing the nucleotide sequence SEQ ID N°4.

A subject of the present invention is also the DNA sequence of the htfIIIA gene as defined above, corresponding to the nucleotide sequence SEQ ID N°3.

The sequence SEQ ID N°3 therefore comprises 1273 nucleotides.

A particular subject of the present invention is the DNA sequence of the htfIIIA gene as defined above, corresponding to the nucleotide sequence SEQ ID N°4. The sequence SEQ ID N°4 therefore comprises 1213 nucleotides.

The sequence SEQ ID N°1 represents the nucleotide sequence of the htFIIIA gene on the upper line according to the present invention i.e. SEQ ID N°3, and the corresponding amino acid sequence (AA) of this nucleotide sequence i.e. SEQ ID N°2 on the lower line.

Figures 1 and 2 below represent the AA sequence coded by htfIIIA of the present invention SEQ ID N°2 on the upper line, and the AA sequences coded by the DREW htfIIIA genes, in Figure 1, and ARAKAWA genes in Figure 2 on the lower line respectively, these DREW and ARAKAWA sequences are as published in the documents referred to above.

Figure 3 below represents the comparison of AA sequences coded by the DREW and ARAKAWA htfIIIA genes respectively with the AA sequence coded by Arakawa htfIIIA on the upper line and the AA sequence coded by DREW htfIIIA on the lower line.

Figure 2 therefore shows, that the corresponding AA

sequence of htfIIIA according to the present invention comprises differences from the AA sequence published in the ARAKAWA article or EP 0704 526, in particular in the corresponding positions 105 and 163, 156 and 214, 320 to 329 and 378 to 387 respectively, these positions being given in relation to the numbering indicated in Figure 2.

Figure 2 also shows that the AA sequence coded by htfIIIA of the present invention begins at position 59 of the AA sequence of Arakawa htfIIIA.

Figure 3 shows that the AA sequences coded by Arakawa and DREW htfIIIA comprise differences at the corresponding positions 214 and 154, 378-387 and 318-327 respectively, these positions being given in relation to the numbering indicated in Figure 3.

Figure 5 shows that the Arakawa htfIIIA sequence codes for a protein, the amino acid sequence of which, indicated in EP 0704 526, begins with the AA methionine specified by the ATG codon which is found in position 20-22 and the translation stops at a TAA codon. If the nucleotide sequence of htfIIIA according to the present invention SEQ ID N°3 is compared with the nucleotide sequence of EP 0704 526 i.e. Arakawa htfIIIA shown in Figure 5 (sequence p11-12-13 of EP 0704 526), it can be noted that it lacks a C nucleotide in position 127 of the EP 0704 526 sequence. This additional C nucleotide results in a shift in the translation of amino acids of this nucleotide sequence: in fact, the ATG which is found in position 20-22 of the ARAKAWA sequence shown in Figure 5 and which is considered to be a start codon of proteinic synthesis by ARAKAWA, is therefore no longer in the same reading frame because of this shift. By taking into consideration this additional C nucleotide, the translation of AA reveals a TGA stop codon in position 57-59 of the ARAKAWA sequence shown in Figure 5. Consequently, the start codon of proteinic synthesis according to the present invention is located downstream of this stop codon. Translation experiments in vitro of SEQ ID N°4 and expression tests in mammalian cells such as Cos cells have made it possible to identify the start codon of the proteinic

synthesis of hTFIIIA according to the present invention.

This start codon of proteinic synthesis of hTFIIIA according to the present invention is the CTG codon in position 176-178 of SEQ ID N°3 (which would correspond to position 194-196 of the ARAKAWA sequence shown in Figure 5).

The coding section of the htFIIIA gene of the present invention begins therefore with this CTG codon which is found in position 176-178 of SEQ ID N°3 which should correspond to the AA Leucine and which in fact corresponds to the AA Methionine as this codon is recognised as a start codon (ref: David S. Peabody The Journal of Biological Chemistry, vol. 264, n°9, pp. 5031-5035, 1989).

Consequently, as Figure 2 shows, the ARAKAWA hTFIIIA protein is longer than the hTFIIIA protein of the present invention.

Furthermore, if the hTFIIIA protein of the present invention and the DREW hTFIIIA protein are compared (comparison shown in Figure 1), it is noticed that the amino acid threonine in position 105 of the hTFIIIA protein of the present invention corresponds to an asparagine residue in position 103 in the DREW hTfIIIA sequence and that the two first AA, M and D of the hTFIIIA protein of the present invention have not been determined for the DREW hTFIIIA protein. The absence of codons specifying these AA and in particular the absence of the start codon of proteinic synthesis, does not permit the expression of this protein. The DREW htfIIIA sequence shown in Figure 4 is therefore incomplete, and this is recognised by the authors of the publication referred to above (DREW et al on page 216 lines 39-41). It can be noted moreover, that the authors of this article also think that the start codon of the DREW htfIIIA sequence should correspond to a methionine coded by ATG as in the ARAKAWA sequence.

The htfIIIA gene according to the present invention is therefore different from the DREW and ARAKAWA htfIIIA genes (EP 0704526) and codes for a hTFIIIA protein, the AA sequence of which is different from that of the DREW and ARAKAWA hTFIIIA proteins.

Therefore a particular subject of the present invention is the DNA sequence of the htfIIIA gene as defined above containing the nucleotide sequence SEQ ID N°3.

A more particular subject of the present invention is the DNA sequence as defined above having the sequence beginning at nucleotide 176 and finishing at nucleotide 1270 of SEQ ID N°3.

One such sequence of the present invention therefore begins at a CTG codon and thus comprises 1095 nucleotides.

A subject of the present invention is the DNA sequence coding for the human transcription factor hTFIIIA as defined above as well as the DNA sequences which hybridize with it and/or show a significant homology with this sequence or fragments of it and coding for proteins having the same function.

By sequences which hybridize are included DNA sequences which hybridize with one of the DNA sequences above under standard conditions of high, medium or low stringency. By proteins with the same function are included polypeptides with the same transcription factor function. The stringency conditions are those carried out in conditions known to a person skilled in the art, such as those described by Sambrook et al (1989) Molecular cloning, Cold Spring Harbor Laboratory Press, 1989. Such stringency conditions are for example hybridization at 65°C, for 18 hours in a 5 x SSPE; 10 x Denhardt's; 100µg/ml ssDNA; 1 % SDS solution followed by washing 3 times for 5 minutes with 2 x SSC; 0.05 % SDS, then washing 3 times for 15 minutes at 65°C in 1 x SSC; 0.1 % SDS. The high stringency conditions for example include hybridization for 18 hours at 65°C in a 5 x SSPE; 10 x Denhardt's; 100µg/ml ssDNA; 1 % SDS solution, followed by washing twice for 20 minutes with a 2 x SSC; 0.05 % SDS solution at 65°C followed by a final wash for 45 minutes in a 0.1 x SSC; 0.1 % SDS solution at 65°C. Medium stringency conditions for example include a final washing for 20 minutes in a 0.2 x SSC, 0.1 % SDS solution at 65°C.

By sequences which show a significant homology are included sequences with a nucleotide sequence with a similarity of at

least 50 % with one of the DNA sequences above and which codes for a protein having the same transcription factor function.

A subject of the present invention is also the DNA sequence as defined above comprising modifications introduced by suppression, insertion and/or substitution of at least one nucleotide coding for a protein having the same biological activity as the human transcription factor htfIIIA.

A particular subject of the present invention is the DNA sequence as defined above as well as similar DNA sequences which have a nucleotide sequence homology of at least 50 % or at least 60 % and preferably at least 70 % with the said DNA sequence.

Therefore a subject of the present invention is also the DNA sequence as defined above as well as the DNA sequences which code for a protein, the AA sequence of which has a homology of at least 40 % and in particular of 45 % or at least 50 %, rather at least 60 % and preferably at least 70 % with the AA sequence coded by the said DNA sequence.

The gene of the present invention is represented as a single strand DNA sequence but it is understood that the present invention includes the complementary DNA sequence of this single strand DNA sequence, and also includes the so-called double strand DNA sequence constituted by these two DNA sequences complementary to each other.

The DNA sequence of the present invention is an example of the combination of codons coding for the amino acids corresponding to the amino acid sequence SEQ ID N°2, but it is also understood that the present invention includes any other arbitrary combination of codons coding for this same amino acid sequence SEQ ID N°2.

The DNA sequence as defined above or this modified DNA sequence as indicated above, can be prepared according to techniques known to a person skilled in the art and in particular those described in the book by Sambrook, J. Fritsh, E. F. § Maniatis, T. (1989) entitled: " Molecular cloning: a laboratory manual ", Laboratory, Cold Spring Harbor NY. In particular the DNA sequence above can be a

cDNA sequence obtained by identification of the 3' and 5' parts of the coding sequence, then amplification of these parts using a DNA polymerase such as pfu polymerase or other DNA polymerases. The introduction, into the oligonucleotide sequence used for PCR, of restriction sites such as Hind III or SmaI allow the cloning of these fragments in appropriate vectors and then the restoration of the sought complete sequence. A detailed description of the operating conditions in which the present invention was carried out is given below.

A quite particular subject of the invention is the polypeptide having the function of human transcription factor hTFIIIA and having the amino acid sequence SEQ ID N°2 coded by the DNA sequence as defined above and the analogues of this polypeptide.

By analogues is understood the polypeptides the amino acid sequence of which has been modified by substitution, suppression or addition of one or more amino acids but which retain the same biological function. Such polypeptide analogues can be produced spontaneously or can be produced by post-transcriptional modification or also by modification of the DNA sequence of the present invention as indicated above, by using techniques known to a person skilled in the art: amongst these techniques, the directed mutagenesis technique (Kramer, W., et al., Nucl. Acids Res., 12, 9441 (1984); Kramer, W. and Fritz, H.J., Methods in Enzymology, 154, 350 (1987); Zoller, M.J. and Smith, M. Methods in Enzymology, 100.468 (1983)) can in particular be mentioned.

Modified DNA synthesis can also be carried out by using well-known chemical synthesis techniques such as the phosphotriester method for example [Letsinger, R.L and Ogilvie, K.K., K. Am. CHEM. Soc., 91.3350 (1969); Merrifield, R.B., Sciences, 150, 178 (1968)] or the phosphoamidite method [Beaucage, S.L and Caruthers, M .H., Tetrahedron Lett., 22, 1859 (1981); McBRIDE, L.J. and Caruthers, M.H. Tetrahedron Lett., 24 245 (1983)] or also by the combination of these methods.

The polypeptides of the present invention can therefore

be prepared by techniques known to a person skilled in the art, in particular partially by chemical synthesis or also by cDNA synthesis by expression in a procaryotic or eucaryotic host cell as indicated below.

A particular subject of the present invention is the process for the preparation of the recombinant htFIIIA protein having the amino acid sequence SEQ ID N°2. This process includes the expression of the DNA sequence as defined above in an appropriate host, then isolation and purification of the said recombinant protein.

To produce the polypeptide of the present invention, recombinant DNA techniques using genetic engineering and cell culture methods known to a person skilled in the art can in particular be used. The following stages can therefore be carried out: firstly preparation of the appropriate gene, then incorporation of this gene into a vector, transfer of the gene carrier vector into an appropriate host cell, expression of the gene and finally purification of the protein coded by this gene.

The DNA sequences according to the present invention and in particular SEQ ID N°3 or SEQ ID N°4 can be prepared according to techniques known to a person skilled in the art, in particular by chemical synthesis, by screening of a gene bank or a cDNA bank using oligonucleotide synthesis probes using known hybridization techniques or also by reverse transcriptase from messenger RNA (mRNA).

The advantage of the technique comprising firstly the isolation of mRNA by extraction of the total RNA then the synthesis of cDNA from this mRNA by reverse transcriptase particularly rests on the fact that the mRNA does not contain introns even though these non-coding sequences are present in the genomic DNA.

The following procedure can in particular be carried out. Firstly the total RNA originating from a cell line such as for example the Raji cell line (RNA Plus, BIOPROBE) is extracted, and from this RNA, synthesis of the sought cDNA is then carried out, in particular by using a kit such as the RNA PCR kit (Perkin Elmer).

It can be noted that within the scope of the present invention, two oligonucleotides located at the ends of the htfIIIA coding sequence published by ARAKAWA (Figure 5) were synthesized i.e. OLT5 and OLT3 and are defined as follows:

- OLT5: 5' CGGGGTACCAAAA ATG CGC AGC AGC GGC GCC GAC 3' i.e. SEQ ID N°5 and
- OLT3: 5' CGGTCTAGA TTA GCC AAG GGT AAG TAC TGC 3' i.e. SEQ ID N°9

but these two oligonucleotides have not made it possible to obtain an amplification product by PCR.

Thus, within the scope of the present invention, the hTFIIIA coding sequence was isolated in two stages: firstly identification of the 3' part then identification of the 5' part.

After identification of the 3' and 5' parts, a HindIII restriction site located on each of these fragments then made it possible to restore the complete sought sequence as indicated below in the experimental part.

The following process was then carried out:

The 3' part was amplified using pfu polymerase (STRATAGENE) using the OLT5.2 and TFIIIA 3'SmaI oligonucleotides as primer i.e.:

- OLT5.2: 5'TCCTTCCCTGACTGCAGCGCC 3' or SEQ ID N°6 and
- TFIIIA3'SmaI: 5'CCT CCC GGG GCC AAG GGT AAG TAC TGC AAC 3' or SEQ ID N°10

The amplification primers are chosen as a function of the part to be amplified according to the usual criteria of a person skilled in the art.

The primers used in the present invention were chosen in the Arakawa htfIIIA sequence shown in Figure 5.

The sequences SEQ ID N°6, SEQ ID N°7 and SEQ ID N°8 are located in positions 320-340 (5'→3'), 361-380 (reverse and complementary sequence) and 391-410 (reverse and complementary sequence) respectively of this Arakawa htfIIIA sequence.

The sequences SEQ ID N°5, SEQ ID N°9 and SEQ ID N°10 are located in positions 20-40 (5'→3'), 1271-1291 (reverse and complementary sequence) and 1268-1288 (reverse and

complementary sequence) respectively of this Arakawa htfIIIA sequence.

It can be noted that sequences SEQ ID N°5, SEQ ID N°9 and SEQ ID N°10 contain sequences corresponding to the restriction enzyme sites i.e. KpnI, XbaI and SmaI respectively.

The oligonucleotide TFIIIA 3' SmaI introduces a restriction site SmaI downstream of the coding sequence. This site permits, if necessary and if required, the fusion of the coding sequence for hTFIIIA with a coding sequence for a hemaglutinin epitope peptide designated " TAG HA ". The expression of the coding sequence for TFIIIA can therefore be combined with that of the coding sequence for TAG HA which can be detected by Western blot analysis, if the fusion gene is expressed.

For identification of the 5' part, this region was isolated by the 5' anchored PCR (5 race System, GIBCO BRL; pfu polymerase, STRATAGENE) technique. Two successive PCR's were carried out using the following oligonucleotides as primer: UAP and TFIIIAPCR5' for the first PCR and UAP and TFIIIA SEQ2 for the second PCR.

UAP is an oligonucleotide provided in the kit.

These oligonucleotides have the following sequences:

- TFIIIAPCR5': 5' CACAAACAAATGGTCTCC 3' or SEQ ID N°8
- TFIIIA SEQ2: 5' TGCACAGGTGCGCGTCAAGC 3' or SEQ ID N°7.

The products of these PCR's i.e. the amplified 5' and 3' fragments are then purified on agarose gel and cloned using the TA cloning kit (INVITROGEN). Sequencing is then carried out: the plasmid DNA of several independent clones is prepared (QIAGEN Plasmids KIT) and the fragments corresponding to the coding sequence of hTFIIIA are sequenced on the two strands (ABI 377XL sequencer, PERKIN ELMER).

The following process can then be carried out according to usual cloning techniques known to a person skilled in the art and in particular cloning by insertion of each fragment into a plasmid provided with the commercial kit (TA cloning Kit Invitrogen), then transformation of a bacterial strain by the plasmid thus obtained is then carried out. The XL1 Blue E. coli strain can in particular be used.

The clones are then cultured in order to extract the plasmid DNA according to standard techniques known to a person skilled in the art referred to above (Sambrook, Fritsh and Maniatis).

Sequencing of the DNA of the amplified fragment contained in the plasmid DNA is carried out.

The compilation of the sequence data thus obtained reveals that in 3', the main part of the isolated sequence corresponds to the htfIIIA sequence of DREW et al. In 5', the longer sequence starts in position 80 of the htfIIIA sequence of Arakawa et al., shown in Figure 2F, and reveals the insertion of a C nucleotide in position 127 in relation to this sequence. If it can be supposed that the synthesis of the cDNA in the application of the technique described above is not complete, the insertion of a nucleotide nevertheless creates a major problem. In fact, the addition of a nucleotide in the coding sequence creates a shift in the reading frame. In order to verify the presence of this nucleotide in the htFIIIA gene, human genomic DNA was analysed by PCR. This DNA was subjected to a PCR reaction using pfu polymerase (STRATAGENE) or Taq polymerase (Perkin Elmer) using the oligonucleotides OLT5 and TFIIIA SEQ2 called SEQ ID N°5 and SEQ ID N°7 respectively as primer. The two PCR products were cloned (TA cloning Kit) then sequenced.

Analysis of the sequence data confirms the presence of this additional C nucleotide in relation to the Arakawa htfIIIA sequence for these two amplifications. The ATG initially described as start codon of proteinic synthesis for Arakawa htfIIIA can therefore no longer be considered as such.

The assembly of 5' and 3' sequences is then carried out and a unique plasmid containing the sought hTFIIIA sequence of the present invention is obtained. The complete hTFIIIA coding sequence is restored in the following manner. A clone originating from the amplification of the genomic DNA is digested using the restriction enzymes EcoRI and HindIII, and after purification, a fragment of approximately 350 bp is obtained. Furthermore, a clone originating from the

amplification of the 3' part using the restriction enzymes HindIII and SmaI is digested and after purification, a fragment of approximately 930 bp is obtained.

The ligation of these fragments in the plasmid pYX223 (expression vector for the yeast - R&D) previously digested by EcoRI and SmaI is then carried out.

A detailed account of the conditions under which the operations indicated above can be carried out is given below in the experimental part. A plasmid is thus obtained in which the gene of the present invention is inserted and this plasmid introduced into a host cell is also thus obtained by operating according to the usual techniques known to a person skilled in the art.

The polypeptide of the present invention can be obtained by expression in a host cell containing the DNA sequence coding for the polypeptide of the invention preceded by a suitable promoter sequence. The host cell can be a procaryotic cell, for example E. coli or a eucaryotic cell such as yeasts, such as for example ascomycetes amongst which are Saccharomyces cerevisiae or also mammalian cells such as Cos. cells

A particular subject of the present invention is an expression vector containing a DNA sequence as defined above.

Thus, such an expression vector according to the present invention contains a DNA sequence which can be the nucleotide sequence SEQ ID N°3 or the sequence beginning at nucleotide 176 and terminating at nucleotide 1270 of SEQ ID N°3. Such an expression vector according to the present invention can also contain the DNA sequences which hybridize with the sequences defined above, and/or show a significant homology with these sequences or fragments of them.

Such an expression vector according to the present invention can also contain DNA sequences which comprise modifications introduced by suppression, insertion and/or substitution of at least one nucleotide coding for a protein with the same biological activity as the human transcription factor hTFIIIA.

Expression vectors are vectors allowing the expression

of the protein under the control of an appropriate promoter. Such a vector can be a plasmid, a cosmid or viral DNA. For the procaryotic cells, the promoter can be for example the lac promoter, trp promoter, tac promoter, β-lactamase promoter or PL promoter. For yeast cells, the promoter can be for example PGK promoter or GAL promoter. For mammalian cells, the promoter can for example be SV40 promoter or adenovirus promoters. Baculovirus type vectors can also be used for expression in insect cells.

The host cells are for example procaryotic cells or eucaryotic cells. The procaryotic cells are for example E. coli, Bacillus or Streptomyces. The eucaryotic host cells comprise yeasts as well as cells of higher organisms, for example mammalian or insect cells. The mammalian cells are for example fibroblasts such as CHO or BHK hamster cells and Cos monkey cells. The insect cells are for example SF9 cells.

The present invention therefore relates to a process which comprises the expression of the htFIIIA protein in a host cell transformed by a DNA coding for the polypeptide sequence corresponding to sequence SEQ ID N°2.

For the implementation of the present invention, the vectors used can for example be pGEX or bpAD and the host cell can be E. coli or for example the vector pYX223, and the host cell can also be S. cerevisiae.

A particular subject of the present invention is a host cell transformed with a vector as defined above, containing the htfIIIA gene according to the present invention.

A very precise subject of the present invention is the plasmid deposited at the CNCM under the number I-2071. It thus concerns the XL1-Blue/bpShtfc2LHA strain containing the htfIIIA gene according to the present invention. The operating conditions in which the present invention was carried out are described below in the experimental part. The hTFIIIA protein coded by the htfIIIA gene is therefore a transcription regulation factor. In fact, the hTFIIIA protein coded by the gene of the present invention has a biological role as a protein binding to the DNA and the

product of this gene is useful as transcription regulation factor.

In particular, the gene of the present invention is expressed in different tissues and probably plays an important role in the initiation of the transcription of the 5S ribosomal RNA gene, and in maintaining the stability of the transcription of other genes in particular involved in control functions. A very large number of diseases accompanying a transcription control disorder have recently been brought to light. It has therefore been noted that certain oncogenic products act as transcription regulation factors and can lead to canceration of cells such as for example in certain leukaemias or also that production of the regulation factor Hox2-4 in too great a quantity induces leukaemia in mice.

Furthermore, in some hereditary diseases, the protein concerned can in itself be normal, the pathogenicity results from the transcription mechanism of the gene coding for this protein. In particular, many hereditary diseases show an abnormality in the quantity of proteins synthesized which is probably due to a disorder in proteinic synthesis which can in particular bring into play the htfIIIA gene and the coded protein as factors involved in the control of the transcription of 5S RNA.

The gene of the present invention can thus be used for the research into abnormalities in the transcription of genes, and in particular in the identification of hereditary diseases for the study of diseases implicating regulation factors and in particular the protein coded by htfIIIA.

The gene of the present invention can also be used for the treatment of certain diseases through transcription control or in the analysis of the pathogenics of these diseases.

The present invention therefore envisages the use of the htfIIIA gene of the present invention and the hTFIIIA protein of the present invention to contribute in particular to the understanding of the transcription mechanism in human beings and also to contribute to the understanding, in the diagnosis and treatment of diseases linked to a disturbance in the

transcription mechanism. Thus hTfIIIA and the htFIIIA protein could be used in the diagnosis or identification of hereditary diseases such as certain cancers or of other diseases resulting from abnormal transcription control. These factors can also be useful in the analysis of the transcription regulation mechanisms.

Therefore a subject of the present invention is the use of the DNA sequence of the gene of the human transcription factor htfIIIA or of the polypeptide having the function of human transcription factor coded by the said DNA sequence as it is defined above, for the preparation of compositions useful in the diagnosis or treatment of diseases linked to a disorder in transcription control.

Such compositions are prepared under the usual conditions known to a person skilled in the art.

A more precise subject of the present invention is the use as defined above in which the disease concerned is cancer. Figures 1 to 5 below show the following illustrations. Figure 1 represents the comparison of the hTFIIIA protein of the present invention with the DREW hTFIIIA protein.

Figure 2 represents the comparison of the hTFIIIA protein of the present invention with the ARAKAWA hTFIIIA protein.

Figure 3 represents the comparison of the DREW hTFIIIA protein with the ARAKAWA hTFIIIA protein.

Figure 4 represents the DREW htfIIIA sequence and the corresponding hTFIIIA protein.

Figure 5 represents the ARAKAWA htfIIIA sequence and the corresponding hTFIIIA protein.

The sequences indicated in the present invention i.e.: SEQ ID N°1 to SEQ ID N°10 are described below.

The experimental part below allows the description of the present invention without however limiting it.

Experimental part

Example 1: cloning and sequencing of the hTFIIIA gene

I) Extraction of total RNA originating from the RAJI human cell line (RNA Plus, BIOPROBE)

The RAJI human cell line was chosen as a source of total RNA.

The RAJI cells used were cultured under the usual culture conditions for this line known to a person skilled in the art.

To extract the total RNA of these cells a standard protocol is carried out using RNA Plus ® (BIOPROBE SYSTEMS) commercial extraction solution.

Then the following is carried out:

a) homogenization:

The cells cultured in suspension are pelleted without being washed beforehand in order to avoid the risk of degradation of the mRNA then are lysed by adding the extraction solution of the RNA Plus ® kit at a rate of 6 ml per 10^7 cells. The samples of homogenate obtained can be stored at - 70 °C.

b) extraction of the RNA:

After homogenization, the homogenate obtained in a) above is left at 4°C for 5 minutes in order to allow the complete disassociation of the nucleoproteinic complexes then 0.2ml of chloroform per 1ml of the RNA Plus ® solution is added, as above in a), the medium is agitated vigorously for 15 seconds and left to rest in ice for 5 minutes, followed by centrifuging at 12000 g and at 4°C, for 15 minutes.

Two clearly visible phases then form: the DNA and the proteins are found in the organic phase (lower phase) and at the interface. The RNA is in the aqueous phase (upper phase) which represents approximately 40 to 50 % of the total volume.

c) Precipitation of the RNA:

The aqueous phase obtained in b) is transferred into a new tube, a volume of isopropanol is added and the sample is placed at 4°C for 15 minutes, followed by centrifuging for 15 minutes at 4°C and at 1200 g. A precipitate is obtained which forms a yellow-white pellet at the bottom of the tube.

d) Washing the RNA:

The supernatant of the solution obtained in c) is eliminated then the pellet is washed with a 75 % ethanol solution using at least 0.8 ml of ethanol per 50 to 100 micrograms of RNA. The medium is mixed (vortex), centrifuged for 10 minutes at 7500 g at 4°C and dried under vacuum. The RNA obtained is

then taken up in 60 microlitres of Tris 10 mM EDTA 1 mM pH=7.5.

II) Synthesis of cDNA

a) Reagents used:

The commercial kit Gene Amp® RNA PCR Kit (Perken Elmer) was used for this cDNA synthesis.

By using this kit, the reverse transcription of RNA to cDNA is firstly obtained by reverse transcriptase MuLV (Murine Leukaemia Virus). An RNase inhibitor isolated from human placenta is included in order to inhibit certain mammalian RNases. The fragments of cDNA are amplified by polymerase chain reaction (PCR). The enzyme used for this reaction is pfu polymerase (Stratagene).

The term dNTP designates the dGTP, dATP, dTTP and dCTP nucleotides.

The term PCR Buffer designates the solution containing 500 mM KCl and 100 mM HCl at pH 8.3.

The term Oligod(T)16 designates a nucleotide sequence constituted by 16 dTTP nucleotides.

Oligonucleotides are used as primers in the technique described below.

The concentrations indicated below represent the final concentrations in the reaction medium.

b) Synthesis of the cDNA by reverse transcription:

2 microlitres of the total RNA (1 microgram) obtained above in 1)d) are pre-incubated at 65°C for 5 minutes, then 8 microlitres of the following reaction solution: 5mM MgCl2, 1xPCR buffer, 1 mM of each dNTP, 5 % of DMSO, 1 U/microlitres of RNase inhibitor, 2.5 U/microlitres of reverse transcriptase MuLV, 2.5 microlitres of oligo(dT)16 is added. The solution is then incubated at 42°C for one hour, then at 99°C for 5 minutes then at 5°C for 5 minutes.

III) Amplification by PCR, cloning and sequencing of the 3' and 5' nucleotide sequences

a) Reaction conditions:

Escherichia coli (E. coli) XL1- Blue type K12 (Stratagene) bacteria was used for the preparation of the plasmids of the present invention.

Growth of this bacteria was carried out according to the usual conditions in LB liquid medium which contains 10 g of bactotryptone, 5 g of yeast extract and 10 g of NaCl per litre of water and which also contains 100 microg/ml of ampicillin (SIGMA).

The colony was removed onto a solid LB + agar + ampicillin medium then cultured in 100 ml of LB medium and incubated to OD (600nm) = 0.8.

The incubation was carried out at 37°C under a normal atmosphere and agitation at 225 rpm.

The viability of the strain is verified when the strain grows on LB + ampicillin medium at 100 microg/ml, the insert containing a gene for resistance to ampicillin bla.

It can be noted that a gene for resistance to ampicillin bla is part of the vector of the kit (TA cloning Kit - Invitrogen) in which the fragments of htfIIIA are cloned. Thus, selection of strains containing the plasmids containing the htfIIIA gene of the present invention can be carried out by culture of the strains in this medium which contains ampicillin (100 microg/ml), such a medium allowing the survival only of strains which contain the gene for resistance to ampicillin and therefore only strains which contain the htfIIIA gene of the present invention.

For the preservation of the strains obtained, 15 % glycerol is added to the culture medium: the cultures are therefore preserved in the LB + 100 micrograms/ml of ampicillin + 15 % of glycerol at the bacterial concentration of OD (600nm) = 0.8 suspension medium in the form of aliquots in cryotubes of 1 ml per tube.

For the sequencing, the plasmid DNA of several bacteria originating from each of the cloning procedures indicated below is prepared using a commercial kit (Qiagen Plasmids kit). The fragments corresponding to the htfIII coding sequence are sequenced on the two strands according to standard techniques known to a person skilled in the art (use of the sequencer ABI 377 XL, Perkin Elmer)

b) Amplification by PCR, and cloning of the 3′ and 5′ nucleotide sequences:

1) Amplification and cloning of the 3' nucleotide sequence
Two amplification primers (primers) were chosen according to the published ARAKAWA HTfIIIA sequence. These OLT3 or TFIIIA3'SmaI and OLT5.2 primers are called SEQ ID N°10 and SEQ ID N°6 respectively.
These oligonucleotides are chosen from the hTFIIIA sequence published by ARAKAWA (Figure 5) and are synthesized according to standard methods known to a person skilled in the art.
The TFIIIA3'SmaI oligonucleotide introduces a restriction site SmaI downstream of the coding sequence. This site will allow the fusion of the htfIIIA 3' nucleotide sequence with a coding sequence for the hemaglutinine TAG peptide.
Thus, the peptide resulting from the expression of the cloned sequence will therefore consist of both the htfIIIA sequence of the present invention and that of TAG HA and can therefore be detected by Western analysis according to usual techniques known to a person skilled in the art.
The following process is then carried out: 2 microlitres of cDNA obtained above in II) b) is added to 50 microlitres of the following reaction solution: 2mM MgCl2, 1xPCR buffer, 200 nanograms/ml of each dNTP, the TFIIIA3'SMAI and OLT5.2 primers at a rate of 0.15 micromoles/l for each, 5 % DMSO and 2.5 U AmpliTaq DNA polymerase.
The cDNA is thus subjected to 30 PCR cycles firstly at 94°C , for one minute then at 65°C for 1 minute then at 72°C for 1 minute.
The products amplified by PCR thus obtained are therefore 3' fragments of approximately 970 base pairs.
The 3' fragments obtained above are cloned in the pCRII vector using the TA cloning Kit (Invitrogen)
The plasmid thus obtained is called 5.2 Raji 2.9.
This plasmid is transferred into the XL1 Blue E. coli strain.
The E. coli strain transformed by the plasmid 5.2 Raji 2.9 is thus obtained.
2) Amplification and cloning of the 5' nucleotide sequence
The 5' portion of the htfIIIA gene of the present invention was isolated using the said 5' anchored PCR technique using a

commercial kit (5'RACE System, Rapid Amplification of cDNA Ends, GIBCO BRL).

Two amplification primers (primers) were chosen from the published ARAKAWA htfIIIA sequence (cf. Figure 5).

These TFIIIAPCR5' and TFIIIA SEQ2 primers are called SEQ ID N°8 and SEQ ID N°7 respectively.

A homopolymeric chain is added to the 3' end of the cDNA using dATP and terminal deoxynucleotidyl transferase (TdT): 10 microlitres of cDNA obtained above in II) b) are incubated at 37°C for 10 minutes in the 1 X tailing buffer reaction solution (Commercial kit solution) and 0.2 mM of dATP and TdT. The TdT is deactivated for 10 minutes at 65°C and the reaction is then brought to 4°C.

The reaction is then directly amplified by PCR: 10 microlitres of the TdT reaction are added to 50 microlitres of PCR reaction solution i.e. 1.5 mM of MgCl2, 1xPCR buffer, 200 nanomoles/ml of each dNTP, UAP and TFIIIA PCR5' primers at a rate of 0.2 micromoles/l for each, 5 % DMSO and 2.5 U AmpliTaq DNApolymerase.

The UAP primer is an oligonucleotide provided with the commercial kit.

The cDNA is thus subjected to 30 PCR cycles, firstly at 94°C for one minute, then at 65°C for 1 minute then at 72°C for 1 minute.

The products amplified by this first PCR i.e. PCR1 are subjected to a second amplification reaction by PCR using the UAP primer and a specific TFIIIASEQ 2 primer. The following process is carried out: 5 microlitres of PCR1 are added to 50 microlitres of the PCR reaction solution indicated below (1.5 mM of MgCl2, 1xPCR buffer, 200 micromoles/l of each dNTP, the UAP and TFIIIA SEQ2 primers at a rate of 0.2 micromoles/l for each, 5 % DMSO and 2.5 U AmpliTaq DNA polymerase.

The DNA is then subjected to 30 PCR cycles, firstly at 94°C for one minute, then at 65°C for 1 minute then at 72°C for 1 minute.

The products amplified by this second PCR i.e. PCR2 are purified on agarose gel. The 5' fragments of approximately 380 base pairs are thus isolated.

The 5' fragments obtained above are thus cloned in the pCRII vector using the TA cloning Kit (Invitrogen).
The plasmid thus obtained is called cDNA-DMSO-3
This plasmid is transferred into the XL1 Blue E. coli strain.
The E. coli strain transformed by the plasmid cDNA-DMSO-3 is thus obtained.

3) Verification of the 5' sequence by amplification of the genomic DNA - Construction of the 5 geno-3 plasmid
Human genomic DNA is extracted from human liver cells according to the usual methods known to a person skilled in the art.
Amplification by PCR of the human genomic DNA is carried out in the following manner:
2 micrograms of human genomic DNA obtained as indicated above is added to 100 microlitres of the following PCR reaction solution: 2mM MgCl2, 1 x native Pfu DNA polymerase buffer, 200 nanograms/ml of each dNTP, the OLT5 and TFIIIA SEQ2 primers at a rate of 0.15 micromoles/l for each, 5 % DMSO and 5 U pfu polymerase.
OLT5 and TFIIIA SEQ2 are called SEQ ID N°5 and SEQ ID N°7 respectively.
The reaction medium is thus subjected to 30 PCR cycles, firstly at 94°C for one minute, then at 60°C for 1 minute, then at 72°C for 1 minute.
The products amplified by PCR thus obtained are fragments of DNA of approximately 360 base pairs.
The fragments thus obtained are cloned in the pCR-Script vector using the pCR-Script SK(+) Cloning kit (Stratagene).
The plasmid thus obtained is called 5 geno-3.
This plasmid is transferred into the XL1 Blue E. coli strain.
The E. coli strain transformed by the plasmid 5 geno-3 is thus obtained.

4) Cloning of the htfIIIA gene according to the present invention.
Construction of the pYX TFIIIALHA plasmid
The complete htfIIIA coding sequence is restored by assembly of the two 3' and 5' fragments obtained above in III) b)1) and III) b)3).

A Hind III restriction site located on each of the 3' and 5' fragments obtained above makes it possible to restore the complete sequence.

The 5 geno-3 plasmid obtained above in III) b)3) is digested by the EcoR1 and HindIII restriction enzymes.

The EcoR1 site is located 11 nucleotides upstream of the coding sequence.

Fragments of approximately 350 base pairs are obtained after purification on agarose gel.

Ligation with the vector pYX/EcoRI + HindIII is then carried out and the vector pYXTFIIIA5' is obtained.

The addition of the 3' fragment to the 5' fragment is then carried out: the 5.2 Raji 2.9 plasmid obtained above in III) b)1), is digested by the restriction enzymes HindIII and SmaI.

After purification on agarose gel, a fragment of approximately 930 base pairs is obtained. This fragment is inserted into the pYXTFIIIA5' plasmid obtained above, previously digested by the restriction enzymes SmaI and HindIII.

The pYXTFIIIALHA plasmid is thus obtained which therefore contains the hTFIIIA gene of the present invention.

Example 2: Construction of the XL1 Blue/pYX TFIIIALHA strain

The preparation of the XL1-Blue/ pYX TFIIIALHA strain, is carried out according to techniques known to a person skilled in the art (ref above: Sambrook, Fritsh and Maniatis) from the XL1- Blue type K12 E. coli strain (Stratagene), and the pYX TFIIIALHA plasmid obtained above in Example 1 is introduced.

Example 3: Construction of the bpS-tfC2LHA plasmid

The vector bpS-SK+ (Stratagene) is used, in which an insert coding for the htFIIIA gene of the present invention is integrated. The following process is carried out: the pYXTFIIIALHA plasmid obtained above in Example 1 is digested by the restriction enzyme EcoRI, this end is filled using DNA Polymerase (Klenow fragment) in the presence of dNTP. This plasmid is then digested by Nhe I and the fragment corresponding to the htfIIIA sequence according to the

present invention is purified. This fragment is inserted into the bpS-SK+ vector prepared as follows: the vector is digested by EcoRI, this site is filled using DNA polymerase then digested by XbaI.

The plasmid bpS-tfC2LHA is thus obtained.

Example 4: Construction of the XL1-Blue/bpS-tfC2LHA strain

For the preparation of the XL1-Blue/bpS-tfC2LHA strain, techniques known to a person skilled in the art, using XL1-Blue type K12 E. coli strain (Stratagene) are carried out, and the bpS-tfC2LHA plasmid obtained above in Example 3 is introduced.

A sample of the strain obtained i.e. XL1- Blue type K12 E. Coli (Stratagene) containing the bpS-SK+ vector (Stratagene) with an insert coding for tfC2 (cDNA coding part containing the htfIIIA coding region) i.e. XL1-Blue/bps-tfC2LHA coding region was deposited at L'Institut Pasteur 25, rue du Docteur ROUX Paris 75015 at the CNCM on the 15th September 1998 under the number I-2071.

Example 5: Identification of the start codon of proteinic synthesis.

Purification of the hTFIII protein was described by Moorefield et al (1994) [reference: the Journal of Biological Chemistry, Vol. 269, N° 33, pp. 20857-20865, 1994, Purification and Characterization of Human Transcription Factor IIIA, B. Moorefield and R. G. Roeder].

The hTFIIIA protein identified by Moorefield has a molecular weight of 42 kDa. It can be noted that the theoretical molecular weight of the htFIIIA protein coded by the Arakawa htfIIIA sequence is 47 kDa.

Proteinic synthesis is generally started at an ATG codon. However the htfIIIA coding sequence of the present invention does not contain the ATG codon in phase.

It has been demonstrated that the different ATG codons, in particular the CTG or GTG codons are start codons of translation in natural cellular transcripts.

With techniques known to a person skilled in the art such as translation experiments in vitro with the htfIIIA sequence according to the invention obtained above in Example 1, and

by expression tests in mammalian cells such as Cos cells, the start codon of hTFIIIA proteinic synthesis according to the present invention was demonstrated.

Within the scope of the present invention, it has thus been demonstrated that the start codon of htfIIIA proteinic synthesis according to the present invention is the CTG codon which is found in position 176-178 of SEQ ID N°3.

Analysis of the results

Analysis of the results obtained by the preparations of the examples indicated above reveal the following points relating to the htfIIIA coding sequence:

- in 3' (above in III) b)1)) the main part of the sequence isolated in the present Application corresponds to the DREW htfIIIA sequence
- in 5' (above in III) b)3)) the longest sequence of fragments obtained by the preparation described above in III) b)3) begins in position 20 of the ARAKAWA htfIIIA sequence and reveals the insertion of a nucleotide in position 127 of the ARAKAWA htfIIIA sequence.

The results obtained by the preparations of htfIIIA described above according to the present invention confirm that omission of a nucleotide in position 127 in the ARAKAWA sequence really does exist in the human htfIIIA gene.

CLAIMS

1) DNA sequence of the htfIIIA gene coding for a protein having the biological function of human transcription factor hTFIIIA.

2) DNA sequence of the htfIIIA gene of the human transcription factor hTFIIIA according to claim 1, coding for the amino acid sequence SEQ ID N°2.

3) DNA sequence of the htfIIIA gene according to claim 1 or 2 containing the nucleotide sequence SEQ ID N°3

4) DNA sequence of the htfIIIA gene according to claims 1 to 3 containing the nucleotide sequence SEQ ID N°4.

5) DNA sequence according to claim 4 having the sequence beginning at nucleotide 176 and finishing at the nucleotide 1270 of SEQ ID N°3.

6) DNA sequence coding for the human transcription factor hTFIIIA according to claims 1 to 5 as well as the DNA sequences which hybridize with it and/or show a significant homology with this sequence or fragments of it and which code for a protein with the same function.

7) DNA sequence according to claims 1 to 6 comprising modifications introduced by suppression, insertion and/or substitution of at least one nucleotide coding for a protein with the same biological activity as human transcription factor hTFIIIA.

8) DNA sequence according to one of claims 1 to 7 as well as similar DNA sequences which have nucleotide sequence homology of at least 50 % or at least 60 % and preferably at least 70 % with the said DNA sequence.

9) DNA sequence according to one of claims 1 to 8 as well as similar DNA sequences which code for a protein, the AA sequence of which has a homology of at least 40 % and in particular 45 % or at least 50 %, rather at least 60 % and preferably at least 70 % with the AA sequence coded by the said DNA sequence.

10) Polypeptide having the function of human transcription factor hTFIIIA and with the amino acid sequence SEQ ID N°2 coded by the DNA sequence according to one of claims 1 to 9 and the analogues of this polypeptide.

11) Process for the preparation of the hTFIIIA recombinant protein having the amino acid sequence SEQ ID N°2 comprising the expression of the DNA sequence according to one of claims 1 to 9 in a appropriate host, then isolation and purification of the said recombinant protein.

12) Expression vector containing the DNA sequence according to one of claims 3 to 9.

13) Host cell transformed with a vector according to claim 12

14) Plasmid deposited at the CNCM under the number I-2071.

15) Use of the human transcription factor htfIIIA gene or of the human transcription factor coded by this gene according to one of the claims 1 to 10 for the preparation of compositions which can be used for the diagnosis or treatment of diseases linked to a disorder in transcription control.

16) Use according to claim 15 for which the disease concerned is cancer.

```
  1 MDPPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSAN 50
    |||||||||||||||||||||||||||||||||||||||||||||||||||
  1 ..PPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSAN 48
            .         .         .         .         .
 51 YSKAWKLDAHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHTGEKPF 100
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 49 YSKAWKLDAHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHTGEKPF 98
            .         .         .         .         .
101 VCAATGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKTFKKHQQLK 150
    |||| |||||||||||||||||||||||||||||||||||||||||||||||
 99 VCAANGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKTFKKHQQLK 148
            .         .         .         .         .
151 IHQCQHTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEGYVCQKGCSFVAK 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
149 IHQCQHTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEGYVCQKGCSFVAK 198
            .         .         .         .         .
201 TWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPERDVCRCPR 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
199 TWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPERDVCRCPR 248
            .         .         .         .         .
251 EGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQSLTRHAVV 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
249 EGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQSLTRHAVV 298
            .         .         .         .         .
301 HDPDKKKMKLKVKKSREKRSLASHLSGYIPPKRKQGQGLSLCQNGESPNC 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
299 HDPDKKKMKLKVKKSREKRSLASHLSGYIPPKRKQGQGLSLCQNGESPNC 348
            .
351 VEDKMLSTVAVLTLG 365
    |||||||||||||||
349 VEDKMLSTVAVLTLG 363
```

FIGURE 1

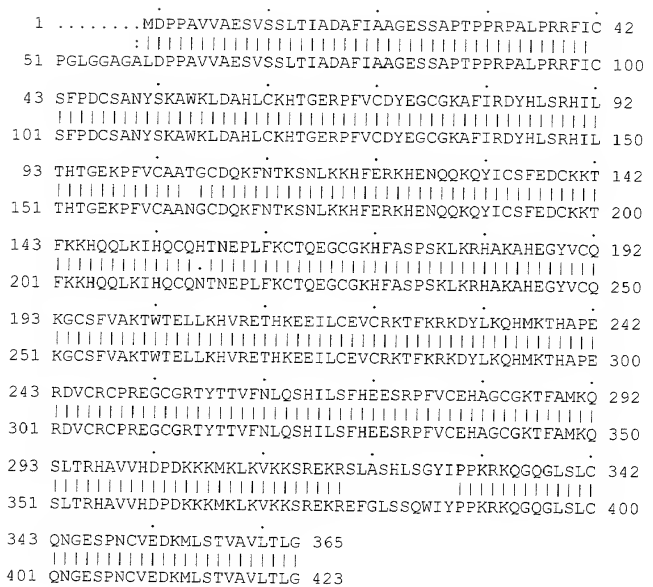

1 .....MDPPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFIC 42
51 PGLGGAGALDPPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFIC 100
43 SFPDCSANYSKAWKLDAHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL 92
101 SFPDCSANYSKAWKLDAHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL 150
93 THTGEKPFVCAATGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKT 142
151 THTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKT 200
143 FKKHQQLKIHQCQHTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEGYVCQ 192
201 FKKHQQLKIHQCQNTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEGYVCQ 250
193 KGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPE 242
251 KGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPE 300
243 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 292
301 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 350
293 SLTRHAVVHDPDKKKMKLKVKKSREKRSLASHLSGYIPPKRKQGQGLSLC 342
351 SLTRHAVVHDPDKKKMKLKVKKSREKREFGLSSQWIYPPKRKQGQGLSLC 400
343 QNGESPNCVEDKMLSTVAVLTLG 365
401 QNGESPNCVEDKMLSTVAVLTLG 423

FIGURE 2

```
 51 PGLGGAGALDPPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFIC 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 ..........PPAVVAESVSSLTIADAFIAAGESSAPTPPRPALPRRFIC  40
    .         .         .         .         .         .
101 SFPDCSANYSKAWKLDAHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 41 SFPDCSANYSKAWKLDAHLCKHTGERPFVCDYEGCGKAFIRDYHLSRHIL  90
    .         .         .         .         .         .
151 THTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKT 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 91 THTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQQKQYICSFEDCKKT 140
    .         .         .         .         .         .
201 FKKHQQLKIHQCQNTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEGYVCQ 250
    |||||||||||||.||||||||||||||||||||||||||||||||||||
141 FKKHQQLKIHQCQHTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEGYVCQ 190
    .         .         .         .         .         .
251 KGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPE 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
191 KGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPE 240
    .         .         .         .         .         .
301 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
241 RDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMKQ 290
    .         .         .         .         .         .
351 SLTRHAVVHDPDKKKMKLKVKKSREKREFGLSSQWIYPPKRKQGQGLSLC 400
    ||||||||||||||||||||||||||||||          ||||||||||||
291 SLTRHAVVHDPDKKKMKLKVKKSREKRSLASHLSGYIPPKRKQGQGLSLC 340
    .         .         .
401 QNGESPNCVEDKMLSTVAVLTLG 423
    |||||||||||||||||||||||
341 QNGESPNCVEDKMLSTVAVLTLG 363
```

FIGURE 3

```
1    CCGCCGGCCGTGGTCGCCGAGTCGGTGTCGTCCTTGACCATCGCCGACGC    50
1    P  P  A  V  V  A  E  S  V  S  S  L  T  I  A  D  A    17

51   GTTCATTGCAGCCGGCGAGAGCTCAGCTCCGACCCCGCCGCGCCCCGCGC   100
18    F  I  A  A  G  E  S  S  A  P  T  P  P  R  P  A  L    34

101  TTCCCAGGAGGTTCATCTGCTCCTTCCCTGACTGCAGCGCCAATTACAGC   150
35     P  R  R  F  I  C  S  F  P  D  C  S  A  N  Y  S      50

151  AAAGCCTGGAAGCTTGACGCGCACCTGTGCAAGCACACGGGGGAGAGACC   200
51   K  A  W  K  L  D  A  H  L  C  K  H  T  G  E  R  P    67

201  ATTTGTTTGTGACTATGAAGGGTGTGGCAAGGCCTTCATCAGGGACTACC   250
68    F  V  C  D  Y  E  G  C  G  K  A  F  I  R  D  Y  H    84

251  ATCTGAGCCGCCACATTCTGACTCACACAGGAGAAAAGCCGTTTGTTTGT   300
85     L  S  R  H  I  L  T  H  T  G  E  K  P  F  V  C      100

301  GCAGCCAATGGCTGTGATCAAAAATTCAACACAAAATCAAACTTGAAGAA   350
101  A  A  N  G  C  D  Q  K  F  N  T  K  S  N  L  K  K    117

351  ACATTTTGAACGCAAACATGAAAATCAACAAAAACAATATATATGCAGTT   400
118   H  F  E  R  K  H  E  N  Q  Q  K  Q  Y  I  C  S  F    134

401  TTGAAGACTGTAAGAAGACCTTTAAGAAACATCAGCAGCTGAAAATCCAT   450
135    E  D  C  K  K  T  F  K  K  H  Q  Q  L  K  I  H      150

451  CAGTGCCAGCATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGGATG   500
151  Q  C  Q  H  T  N  E  P  L  F  K  C  T  Q  E  G  C    167

501  TGGGAAACACTTTGCATCACCCAGCAAGCTGAAACGACATGCCAAGGCCC   550
168   G  K  H  F  A  S  P  S  K  L  K  R  H  A  K  A  H    184

551  ACGAGGGCTATGTATGTCAAAAAGGATGTTCCTTTGTGGCAAAAACATGG   600
185    E  G  Y  V  C  Q  K  G  C  S  F  V  A  K  T  W      200

601  ACGGAACTTCTGAAACATGTGAGAGAAACCCATAAAGAGGAAATACTATG   650
201  T  E  L  L  K  H  V  R  E  T  H  K  E  E  I  L  C    217
```

FIGURE 4

```
651  TGAAGTATGCCGGAAAACATTTAAACGCAAAGATTACCTTAAGCAACACA  700
218   E  V  C  R  K  T  F  K  R  K  D  Y  L  K  Q  H  M  234

701  TGAAAACTCATGCCCCAGAAAGGGATGTATGTCGCTGTCCAAGAGAAGGC  750
235    K  T  H  A  P  E  R  D  V  C  R  C  P  R  E  G    250

751  TGTGGAAGAACCTATACAACTGTGTTTAATCTCCAAAGCCATATCCTCTC  800
251   C  G  R  T  Y  T  T  V  F  N  L  Q  S  H  I  L  S  267

801  CTTCCATGAGGAAAGCCGCCCTTTTGTGTGTGAACATGCTGGCTGTGGCA  850
268   F  H  E  E  S  R  P  F  V  C  E  H  A  G  C  G  K  284

851  AAACATTTGCAATGAAACAAAGTCTCACTAGGCATGCTGTTGTACATGAT  900
285    T  F  A  M  K  Q  S  L  T  R  H  A  V  V  H  D    300

901  CCTGACAAGAAGAAAATGAAGCTCAAAGTCAAAAAATCTCGTGAAAAACG  950
301   P  D  K  K  K  M  K  L  K  V  K  K  S  R  E  K  R  317

951  GAGTTTGGCCTCTCATCTCAGTGGATATATCCCTCCCAAAAGGAAACAAG  1000
318   S  L  A  S  H  L  S  G  Y  I  P  P  K  R  K  Q  G  334

1001 GGCAAGGCTTATCTTTGTGTCAAAACGGAGAGTCACCCAACTGTGTGGAA  1050
335    Q  G  L  S  L  C  Q  N  G  E  S  P  N  C  V  E    350

1051 GACAAGATGCTCTCGACAGTTGCAGTACTTACCCTTGGCTAAGAACTGCA  1100
351   D  K  M  L  S  T  V  A  V  L  T  L  G  *           364

1101 CTGCTTTGTTTAAAGGACTGCAGACCAAGGAGCGAGCTTTCTCTCAGAGC  1150

1151   ATGCTTTTCTTTATTAAAATTAC  1173
```

FIGURE 4

```
1    ATGCGCGATCTCCCGGAGCATGCGCAGCAGCGGCGCCGACGCGGGGCGGT   50
1                        M  R  S  S  G  A  D  A  G  R  C   11

51   GCCTGGTGACCGCGCGCGCTCCCGGAAGTGTGCCGGCGTCGCGCGAAGGT  100
12     L  V  T  A  R  A  P  G  S  V  P  A  S  R  E  G    27

101  TCAGCAGGGAGCCGTGGGCCGGGCGCGCGGTTCCCGGCACGTGTCTCGGC  150
28   S  A  G  S  R  G  P  G  A  R  F  P  A  R  V  S  A   44

151  ACGTGGCAGCGCGCCTGGCCCTGGGCTTGGAGGCGCCGGCGCCCTGGATC  200
45    R  G  S  A  P  G  P  G  L  G  G  A  G  A  L  D  P   61

201  CGCCGGCCGTGGTCGCCGAGTCGGTGTCGTCCTTGACCATCGCCGACGCG  250
62     P  A  V  V  A  E  S  V  S  S  L  T  I  A  D  A    77

251  TTCATTGCAGCCGGCGAGAGCTCAGCTCCGACCCCGCCGCGCCCCGCGCT  300
78   F  I  A  A  G  E  S  S  A  P  T  P  P  R  P  A  L   94

301  TCCCAGGAGGTTCATCTGCTCCTTCCCTGACTGCAGCGCCAATTACAGCA  350
95    P  R  R  F  I  C  S  F  P  D  C  S  A  N  Y  S  K  111

351  AAGCCTGGAAGCTTGACGCGCACCTGTGCAAGCACACGGGGGAGAGACCA  400
112    A  W  K  L  D  A  H  L  C  K  H  T  G  E  R  P    127

401  TTTGTTTGTGACTATGAAGGGTGTGGCAAGGCCTTCATCAGGGACTACCA  450
128  F  V  C  D  Y  E  G  C  G  K  A  F  I  R  D  Y  H   144

451  TCTGAGCCGCCACATTCTGACTCACACAGGAGAAAAGCCGTTTGTTTGTG  500
145   L  S  R  H  I  L  T  H  T  G  E  K  P  F  V  C  A  161

501  CAGCCAATGGCTGTGATCAAAAATTCAACACAAAATCAAACTTGAAGAAA  550
162    A  N  G  C  D  Q  K  F  N  T  K  S  N  L  K  K   177

551  CATTTTGAACGCAAACATGAAAATCAACAAAAACAATATATATGCAGTTT  600
178  H  F  E  R  K  H  E  N  Q  Q  K  Q  Y  I  C  S  F   194

601  TGAAGACTGTAAGAAGACCTTTAAGAAACATCAGCAGCTGAAAATCCATC  650
195   E  D  C  K  K  T  F  K  K  H  Q  Q  L  K  I  H  Q  211

651  AGTGCCAGAATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGGATGT  700
212    C  Q  N  T  N  E  P  L  F  K  C  T  Q  E  G  C   227
```

FIGURE 5

```
701  GGGAAACACTTTGCATCACCCAGCAAGCTGAAACGACATGCCAAGGCCCA  750
228  G  K  H  F  A  S  P  S  K  L  K  R  H  A  K  A  H   244

751  CGAGGGCTATGTATGTCAAAAAGGATGTTCCTTTGTGGCAAAAACATGGA  800
245   E  G  Y  V  C  Q  K  G  C  S  F  V  A  K  T  W  T  261

801  CGGAACTTCTGAAACATGTGAGAGAAACCCATAAAGAGGAAATACTATGT  850
262    E  L  L  K  H  V  R  E  T  H  K  E  E  I  L  C   277

851  GAAGTATGCCGGAAAACATTTAAACGCAAAGATTACCTTAAGCAACACAT  900
278  E  V  C  R  K  T  F  K  R  K  D  Y  L  K  Q  H  M  294

901  GAAAACTCATGCCCCAGAAAGGGATGTATGTCGCTGTCCAAGAGAAGGCT  950
295   K  T  H  A  P  E  R  D  V  C  R  C  P  R  E  G  C  311

951  GTGGAAGAACCTATACAACTGTGTTTAATCTCCAAAGCCATATCCTCTCC  1000
312    G  R  T  Y  T  T  V  F  N  L  Q  S  H  I  L  S   327

1001 TTCCATGAGGAAAGCCGCCCTTTTGTGTGTGAACATGCTGGCTGTGGCAA  1050
328  F  H  E  E  S  R  P  F  V  C  E  H  A  G  C  G  K  344

1051 AACATTTGCAATGAAACAAAGTCTCACTAGGCATGCTGTTGTACATGATC  1100
345   T  F  A  M  K  Q  S  L  T  R  H  A  V  V  H  D  P  361

1101 CTGACAAGAAGAAAATGAAGCTCAAAGTCAAAAAATCTCGTGAAAAACGG  1150
362    D  K  K  K  M  K  L  K  V  K  K  S  R  E  K  R   377

1151 GAGTTTGGCCTCTCATCTCAGTGGATATATCCTCCCAAAAGGAAACAAGG  1200
378  E  F  G  L  S  S  Q  W  I  Y  P  P  K  R  K  Q  G  394

1201 GCAAGGCTTATCTTTGTGTCAAAACGGAGAGTCACCCAACTGTGTGGAAG  1250
395   Q  G  L  S  L  C  Q  N  G  E  S  P  N  C  V  E  D  411

1251 ACAAGATGCTCTCGACAGTTGCAGTACTTACCCTTGGCTAAGAACTGCAC  1300
412    K  M  L  S  T  V  A  V  L  T  L  G  *              424

1301 TGCTTTGTTTAAAGGACTGCAGACCAAGGAGTCGAGCTTTCTCTCAGAGC  1350

ATGCTTTTCTTTATTAAAATTACTGATGCAGAAAAAAAAAAAAAAAAAA     1399
```

FIGURE 5

Our Ref.: 146.1364

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:
F. BORDON-PALLIER et al
PCT/FR99/02738 : PCT Date: November 9, 1999
Serial No.:
Filed: Concurrently Herewith
For: HUMAN...PROTEIN

600 Third Avenue
New York, NY 10016
May 8, 2001

RE: SEQUENCE LISTING

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Applicants are submitting herewith a paper copy of the Sequence Listing and a diskette corresponding thereto.

Respectfully submitted,
BIERMAN, MUSERLIAN AND LUCAS

Charles A. Muserlian, #19,683
Attorney for Applicant(s)
Tel. # (212) 661-8000

CAM:sd
Enclosures: Paper Copy of Sequence Listing
Diskette
Return Receipt Postcard

SEQUENCE LISTING

<110> Hoechst Marion Roussel

<120> HUMAN htFIII GENE AND CODED htfIIIA PROTEIN

<130> 9823seq

<140>
<141>

<160> 10

<170> PatentIn Vers. 2.0

<210> 1
<211> 1273
<212> ADN
<213> Human

<220>
<221> CDS
<222> (176)..(1270)

<400> SEQ ID No: 1

```
atgcgcagca gcggcgccga cgcggggcgg tgcctggtga ccgcgcgcgc tcccggaagt 60

gtgccggcgt cgcgcgaagg ttcagcaggg agccgtgggc cgggcgcgcc ggttcccggc 120

acgtgtctcg gcacgtggca gcgcgcctgg ccctgggctt ggaggcgccg gcgcc ctg  178
                                                            Leu
                                                             1

gat ccg ccg gcc gtg gtc gcc gag tcg gtg tcg tcc ttg acc atc gcc   226
Asp Pro Pro Ala Val Val Ala Glu Ser Val Ser Ser Leu Thr Ile Ala
                5                  10                  15

gac gcg ttc att gca gcc ggc gag agc tca gct ccg acc ccg ccg cgc   274
Asp Ala Phe Ile Ala Ala Gly Glu Ser Ser Ala Pro Thr Pro Pro Arg
        20                  25                  30

ccc gcg ctt ccc agg agg ttc atc tgc tcc ttc cct gac tgc agc gcc   322
Pro Ala Leu Pro Arg Arg Phe Ile Cys Ser Phe Pro Asp Cys Ser Ala
    35                  40                  45

aat tac agc aaa gcc tgg aag ctt gac gcg cac ctg tgc aag cac acg   370
Asn Tyr Ser Lys Ala Trp Lys Leu Asp Ala His Leu Cys Lys His Thr
 50                  55                  60                  65

ggg gag aga cca ttt gtt tgt gac tat gaa ggg tgt ggc aag gcc ttc   418
Gly Glu Arg Pro Phe Val Cys Asp Tyr Glu Gly Cys Gly Lys Ala Phe
                70                  75                  80

atc agg gac tac cat ctg agc cgc cac att ctg act cac aca gga gaa   466
Ile Arg Asp Tyr His Leu Ser Arg His Ile Leu Thr His Thr Gly Glu
            85                  90                  95
```

```
aag ccg ttt gtt tgt gca gcc act ggc tgt gat caa aaa ttc aac aca   514
Lys Pro Phe Val Cys Ala Ala Thr Gly Cys Asp Gln Lys Phe Asn Thr
        100                 105                 110

aaa tca aac ttg aag aaa cat ttt gaa cgc aaa cat gaa aat caa caa   562
Lys Ser Asn Leu Lys Lys His Phe Glu Arg Lys His Glu Asn Gln Gln
    115                 120                 125

aaa caa tat ata tgc agt ttt gaa gac tgt aag aag acc ttt aag aaa   610
Lys Gln Tyr Ile Cys Ser Phe Glu Asp Cys Lys Lys Thr Phe Lys Lys
130                 135                 140                 145

cat cag cag ctg aaa atc cat cag tgc cag cat acc aat gaa cct cta   658
His Gln Gln Leu Lys Ile His Gln Cys Gln His Thr Asn Glu Pro Leu
                150                 155                 160

ttc aag tgt acc cag gaa gga tgt ggg aaa cac ttt gca tca ccc agc   706
Phe Lys Cys Thr Gln Glu Gly Cys Gly Lys His Phe Ala Ser Pro Ser
            165                 170                 175

aag ctg aaa cga cat gcc aag gcc cac gag ggc tat gta tgt caa aaa   754
Lys Leu Lys Arg His Ala Lys Ala His Glu Gly Tyr Val Cys Gln Lys
        180                 185                 190

gga tgt tcc ttt gtg gca aaa aca tgg acg gaa ctt ctg aaa cat gtg   802
Gly Cys Ser Phe Val Ala Lys Thr Trp Thr Glu Leu Leu Lys His Val
    195                 200                 205

aga gaa acc cat aaa gag gaa ata cta tgt gaa gta tgc cgg aaa aca   850
Arg Glu Thr His Lys Glu Glu Ile Leu Cys Glu Val Cys Arg Lys Thr
210                 215                 220                 225

ttt aaa cgc aaa gat tac ctt aag caa cac atg aaa act cat gcc cca   898
Phe Lys Arg Lys Asp Tyr Leu Lys Gln His Met Lys Thr His Ala Pro
                230                 235                 240

gaa agg gat gta tgt cgc tgt cca aga gaa ggc tgt gga aga acc tat   946
Glu Arg Asp Val Cys Arg Cys Pro Arg Glu Gly Cys Gly Arg Thr Tyr
            245                 250                 255

act act gtg ttt aat ctc caa agc cat atc ctc tcc ttc cat gag gaa   994
Thr Thr Val Phe Asn Leu Gln Ser His Ile Leu Ser Phe His Glu Glu
        260                 265                 270

agc cgc cct ttt gtg tgt gaa cat gct ggc tgt ggc aaa aca ttt gca  1042
Ser Arg Pro Phe Val Cys Glu His Ala Gly Cys Gly Lys Thr Phe Ala
    275                 280                 285

atg aaa caa agt ctc act agg cat gct gtt gta cat gat cct gac aag  1090
Met Lys Gln Ser Leu Thr Arg His Ala Val Val His Asp Pro Asp Lys
290                 295                 300                 305

aag aaa atg aag ctc aaa gtc aaa aaa tct cgt gaa aaa cgg agt ttg  1138
Lys Lys Met Lys Leu Lys Val Lys Lys Ser Arg Glu Lys Arg Ser Leu
                310                 315                 320

gcc tct cat ctc agt gga tat atc cct ccc aaa agg aaa caa ggg caa  1186
```

```
Ala Ser His Leu Ser Gly Tyr Ile Pro Pro Lys Arg Lys Gln Gly Gln
            325                 330                 335

ggc tta tct ttg tgt caa aac gga gag tca ccc aac tgt gtg gaa gac   1234
Gly Leu Ser Leu Cys Gln Asn Gly Glu Ser Pro Asn Cys Val Glu Asp
        340                 345                 350

aag atg ctc tcg aca gtt gca gta ctt acc ctt ggc taa               1273
Lys Met Leu Ser Thr Val Ala Val Leu Thr Leu Gly
    355                 360                 365


<210> 2
<211> 365
<212> PRT
<213> Human

<400> SEQ ID No: 2
Leu Asp Pro Pro Ala Val Val Ala Glu Ser Val Ser Ser Leu Thr Ile
  1               5                  10                  15

Ala Asp Ala Phe Ile Ala Ala Gly Glu Ser Ser Ala Pro Thr Pro Pro
             20                  25                  30

Arg Pro Ala Leu Pro Arg Arg Phe Ile Cys Ser Phe Pro Asp Cys Ser
         35                  40                  45

Ala Asn Tyr Ser Lys Ala Trp Lys Leu Asp Ala His Leu Cys Lys His
     50                  55                  60

Thr Gly Glu Arg Pro Phe Val Cys Asp Tyr Glu Gly Cys Gly Lys Ala
 65                  70                  75                  80

Phe Ile Arg Asp Tyr His Leu Ser Arg His Ile Leu Thr His Thr Gly
                 85                  90                  95

Glu Lys Pro Phe Val Cys Ala Ala Thr Gly Cys Asp Gln Lys Phe Asn
            100                 105                 110

Thr Lys Ser Asn Leu Lys Lys His Phe Glu Arg Lys His Glu Asn Gln
        115                 120                 125

Gln Lys Gln Tyr Ile Cys Ser Phe Glu Asp Cys Lys Lys Thr Phe Lys
    130                 135                 140

Lys His Gln Gln Leu Lys Ile His Gln Cys Gln His Thr Asn Glu Pro
145                 150                 155                 160

Leu Phe Lys Cys Thr Gln Glu Gly Cys Gly Lys His Phe Ala Ser Pro
                165                 170                 175

Ser Lys Leu Lys Arg His Ala Lys Ala His Glu Gly Tyr Val Cys Gln
            180                 185                 190

Lys Gly Cys Ser Phe Val Ala Lys Thr Trp Thr Glu Leu Leu Lys His
        195                 200                 205
```

Val Arg Glu Thr His Lys Glu Glu Ile Leu Cys Glu Val Cys Arg Lys
210 215 220

Thr Phe Lys Arg Lys Asp Tyr Leu Lys Gln His Met Lys Thr His Ala
225 230 235 240

Pro Glu Arg Asp Val Cys Arg Cys Pro Arg Glu Gly Cys Gly Arg Thr
245 250 255

Tyr Thr Thr Val Phe Asn Leu Gln Ser His Ile Leu Ser Phe His Glu
260 265 270

Glu Ser Arg Pro Phe Val Cys Glu His Ala Gly Cys Gly Lys Thr Phe
275 280 285

Ala Met Lys Gln Ser Leu Thr Arg His Ala Val Val His Asp Pro Asp
290 295 300

Lys Lys Lys Met Lys Leu Lys Val Lys Lys Ser Arg Glu Lys Arg Ser
305 310 315 320

Leu Ala Ser His Leu Ser Gly Tyr Ile Pro Pro Lys Arg Lys Gln Gly
325 330 335

Gln Gly Leu Ser Leu Cys Gln Asn Gly Glu Ser Pro Asn Cys Val Glu
340 345 350

Asp Lys Met Leu Ser Thr Val Ala Val Leu Thr Leu Gly
355 360 365

<210> 3
<211> 1273
<212> ADN
<213> Human

<400> SEQ ID No: 3
atgcgcagca gcggcgccga cgcggggcgg tgcctggtga ccgcgcgcgc tcccggaagt 60

gtgccggcgt cgcgcgaagg ttcagcaggg agccgtgggc cgggcgcgcc ggttcccggc 120

acgtgtctcg gcacgtggca gcgcgcctgg ccctgggctt ggaggcgccg gcgccctgga 180

tccgccggcc gtggtcgccg agtcggtgtc gtccttgacc atcgccgacg cgttcattgc 240

agccggcgag agctcagctc cgaccccgcc gcgccccgcg cttcccagga ggttcatctg 300

ctccttccct gactgcagcg ccaattacag caaagcctgg aagcttgacg cgcacctgtg 360

caagcacacg ggggagagac catttgtttg tgactatgaa gggtgtggca aggccttcat 420

cagggactac catctgagcc gccacattct gactcacaca ggagaaaagc cgtttgtttg 480

tgcagccact ggctgtgatc aaaaattcaa cacaaaatca aacttgaaga aacattttga 540

acgcaaacat gaaaatcaac aaaaacaata tatatgcagt tttgaagact gtaagaagac 600

ctttaagaaa catcagcagc tgaaaatcca tcagtgccag cataccaatg aacctctatt 660

caagtgtacc caggaaggat gtgggaaaca ctttgcatca cccagcaagc tgaaacgaca 720

tgccaaggcc cacgagggct atgtatgtca aaaaggatgt tcctttgtgg caaaaacatg 780

gacggaactt ctgaaacatg tgagagaaac ccataaagag gaaatactat gtgaagtatg 840

ccggaaaaca tttaaacgca aagattacct taagcaacac atgaaaactc atgccccaga 900

aagggatgta tgtcgctgtc caagagaagg ctgtggaaga acctatacta ctgtgtttaa 960

tctccaaagc catatcctct ccttccatga ggaaagccgc ccttttgtgt gtgaacatgc 1020

tggctgtggc aaaacatttg caatgaaaca aagtctcact aggcatgctg ttgtacatga 1080

tcctgacaag aagaaaatga agctcaaagt caaaaaatct cgtgaaaaac ggagtttggc 1140

ctctcatctc agtggatata tccctcccaa aaggaaacaa gggcaaggct tatctttgtg 1200

tcaaaacgga gagtcaccca actgtgtgga agacaagatg ctctcgacag ttgcagtact 1260

taccccttggc taa 1273

<210> 4
<211> 1213
<212> ADN
<213> Human

<400> SEQ ID No: 4
gtgccggcgc cgcgcgaagg ttcagcaggg agccgtgggc cgggcgcgcc ggttcccggc 60

acgtgtctcg gcacgtggca gcgcgcctgg ccctgggctt ggaggcgccg gcgccctgga 120

tccgccggcc gtggtcgccg agtcggtgtc gtccttgacc atcgccgacg cgttcattgc 180

agccggcgag agctcagctc cgaccccgcc gcgccccgcg cttcccagga ggttcatctg 240

ctccttccct gactgcagcg ccaattacag caaagcctgg aagcttgacg cgcacctgtg 300

caagcacacg ggggagagac catttgtttg tgactatgaa gggtgtggca aggccttcat 360

cagggactac catctgagcc gccacattct gactcacaca ggagaaaagc cgtttgtttg 420

tgcagccact ggctgtgatc aaaaattcaa cacaaaatca aacttgaaga aacattttga 480

acgcaaacat gaaaatcaac aaaaacaata tatatgcagt tttgaagact gtaagaagac 540

ctttaagaaa catcagcagc tgaaaatcca tcagtgccag cataccaatg aacctctatt 600

caagtgtacc caggaaggat gtgggaaaca ctttgcatca cccagcaagc tgaaacgaca 660

tgccaaggcc cacgagggct atgtatgtca aaaaggatgt tcctttgtgg caaaaacatg 720

gacggaactt ctgaaacatg tgagagaaac ccataaagag gaaatactat gtgaagtatg 780

ccggaaaaca tttaaacgca aagattacct taagcaacac atgaaaactc atgccccaga 840

aagggatgta tgtcgctgtc caagagaagg ctgtggaaga acctatacta ctgtgtttaa 900

tctccaaagc catatcctct ccttccatga ggaaagccgc ccttttgtgt gtgaacatgc 960

tggctgtggc aaaacatttg caatgaaaca aagtctcact aggcatgctg ttgtacatga 1020

tcctgacaag aagaaaatga agctcaaagt caaaaaatct cgtgaaaaac ggagtttggc 1080

ctctcatctc agtggatata tccctcccaa aaggaaacaa gggcaaggct tatctttgtg 1140

tcaaaacgga gagtcaccca actgtgtgga agacaagatg ctctcgacag ttgcagtact 1200

taccccttggc taa 1213

<210> 5
<211> 34
<212> ADN
<213> Human

<400> SEQ ID No: 5
cggggtacca aaaatgcgca gcagcggcgc cgac 34

<210> 6
<211> 21
<212> ADN
<213> Human

<400> SEQ ID No: 6
tccttccctg actgcagcgc c 21

<210> 7
<211> 20
<212> ADN
<213> Human

<400> SEQ ID No: 7
tgcacaggtg cgcgtcaagc 20

<210> 8
<211> 20
<212> ADN
<213> Human

<400> SEQ ID No: 8
cacaaacaaa tggtctctcc 20

<210> 9
<211> 30
<212> ADN
<213> Human

<400> SEQ ID No: 9
cggtctagat tagccaaggg taagtactgc 30

<210> 10
<211> 30
<212> ADN
<213> Human

<400> SEQ ID No: 10
cctcccgggg ccaagggtaa gtactgcaac 30

SEQUENCE LISTING

<110> Hoechst Marion Roussel

<120> Human htFIIIA gene and coded htfIIIA protein

<130> 9823seq

<140>
<141>

<160> 10

<170> PatentIn Vers. 2.0

<210> 1
<211> 1273
<212> DNA
<213> Human

<220>
<221> CDS
<222> (176)..(1270)

<400> 1
atgcgcagca gcggcgccga cgcggggcgg tgcctggtga ccgcgcgcgc tcccggaagt 60

gtgccggcgt cgcgcgaagg ttcagcaggg agccgtgggc cgggcgcgcc ggttcccggc 120

acgtgtctcg gcacgtggca gcgcgcctgg ccctgggctt ggaggcgccg gcgcc ctg 178
Met
1

gat ccg ccg gcc gtg gtc gcc gag tcg gtg tcg tcc ttg acc atc gcc 226
Asp Pro Pro Ala Val Val Ala Glu Ser Val Ser Ser Leu Thr Ile Ala
5 10 15

gac gcg ttc att gca gcc ggc gag agc tca gct ccg acc ccg ccg cgc 274
Asp Ala Phe Ile Ala Ala Gly Glu Ser Ser Ala Pro Thr Pro Pro Arg
20 25 30

ccc gcg ctt ccc agg agg ttc atc tgc tcc ttc cct gac tgc agc gcc 322
Pro Ala Leu Pro Arg Arg Phe Ile Cys Ser Phe Pro Asp Cys Ser Ala
35 40 45

aat tac agc aaa gcc tgg aag ctt gac gcg cac ctg tgc aag cac acg 370
Asn Tyr Ser Lys Ala Trp Lys Leu Asp Ala His Leu Cys Lys His Thr
50 55 60 65

ggg gag aga cca ttt gtt tgt gac tat gaa ggg tgt ggc aag gcc ttc 418
Gly Glu Arg Pro Phe Val Cys Asp Tyr Glu Gly Cys Gly Lys Ala Phe
70 75 80

atc agg gac tac cat ctg agc cgc cac att ctg act cac aca gga gaa 466
Ile Arg Asp Tyr His Leu Ser Arg His Ile Leu Thr His Thr Gly Glu
85 90 95

aag ccg ttt gtt tgt gca gcc act ggc tgt gat caa aaa ttc aac aca 514
Lys Pro Phe Val Cys Ala Ala Thr Gly Cys Asp Gln Lys Phe Asn Thr
100 105 110

```
aaa tca aac ttg aag aaa cat ttt gaa cgc aaa cat gaa aat caa caa   562
Lys Ser Asn Leu Lys Lys His Phe Glu Arg Lys His Glu Asn Gln Gln
    115                 120                 125

aaa caa tat ata tgc agt ttt gaa gac tgt aag aag acc ttt aag aaa   610
Lys Gln Tyr Ile Cys Ser Phe Glu Asp Cys Lys Lys Thr Phe Lys Lys
130                 135                 140                 145

cat cag cag ctg aaa atc cat cag tgc cag cat acc aat gaa cct cta   658
His Gln Gln Leu Lys Ile His Gln Cys Gln His Thr Asn Glu Pro Leu
                150                 155                 160

ttc aag tgt acc cag gaa gga tgt ggg aaa cac ttt gca tca ccc agc   706
Phe Lys Cys Thr Gln Glu Gly Cys Gly Lys His Phe Ala Ser Pro Ser
            165                 170                 175

aag ctg aaa cga cat gcc aag gcc cac gag ggc tat gta tgt caa aaa   754
Lys Leu Lys Arg His Ala Lys Ala His Glu Gly Tyr Val Cys Gln Lys
        180                 185                 190

gga tgt tcc ttt gtg gca aaa aca tgg acg gaa ctt ctg aaa cat gtg   802
Gly Cys Ser Phe Val Ala Lys Thr Trp Thr Glu Leu Leu Lys His Val
    195                 200                 205

aga gaa acc cat aaa gag gaa ata cta tgt gaa gta tgc cgg aaa aca   850
Arg Glu Thr His Lys Glu Glu Ile Leu Cys Glu Val Cys Arg Lys Thr
210                 215                 220                 225

ttt aaa cgc aaa gat tac ctt aag caa cac atg aaa act cat gcc cca   898
Phe Lys Arg Lys Asp Tyr Leu Lys Gln His Met Lys Thr His Ala Pro
                230                 235                 240

gaa agg gat gta tgt cgc tgt cca aga gaa ggc tgt gga aga acc tat   946
Glu Arg Asp Val Cys Arg Cys Pro Arg Glu Gly Cys Gly Arg Thr Tyr
            245                 250                 255

act act gtg ttt aat ctc caa agc cat atc ctc tcc ttc cat gag gaa   994
Thr Thr Val Phe Asn Leu Gln Ser His Ile Leu Ser Phe His Glu Glu
        260                 265                 270

agc cgc cct ttt gtg tgt gaa cat gct ggc tgt ggc aaa aca ttt gca  1042
Ser Arg Pro Phe Val Cys Glu His Ala Gly Cys Gly Lys Thr Phe Ala
    275                 280                 285

atg aaa caa agt ctc act agg cat gct gtt gta cat gat cct gac aag  1090
Met Lys Gln Ser Leu Thr Arg His Ala Val Val His Asp Pro Asp Lys
290                 295                 300                 305

aag aaa atg aag ctc aaa gtc aaa aaa tct cgt gaa aaa cgg agt ttg  1138
Lys Lys Met Lys Leu Lys Val Lys Lys Ser Arg Glu Lys Arg Ser Leu
                310                 315                 320

gcc tct cat ctc agt gga tat atc cct ccc aaa agg aaa caa ggg caa  1186
Ala Ser His Leu Ser Gly Tyr Ile Pro Pro Lys Arg Lys Gln Gly Gln
            325                 330                 335

ggc tta tct ttg tgt caa aac gga gag tca ccc aac tgt gtg gaa gac  1234
Gly Leu Ser Leu Cys Gln Asn Gly Glu Ser Pro Asn Cys Val Glu Asp
        340                 345                 350

aag atg ctc tcg aca gtt gca gta ctt acc ctt ggc taa              1273
Lys Met Leu Ser Thr Val Ala Val Leu Thr Leu Gly
    355                 360                 365
```

<210> 2

<211> 365
<212> PRT
<213> Human

<400> 2

```
Met Asp Pro Pro Ala Val Val Ala Glu Ser Val Ser Ser Leu Thr Ile
  1               5                  10                  15

Ala Asp Ala Phe Ile Ala Ala Gly Glu Ser Ser Ala Pro Thr Pro Pro
             20                  25                  30

Arg Pro Ala Leu Pro Arg Arg Phe Ile Cys Ser Phe Pro Asp Cys Ser
         35                  40                  45

Ala Asn Tyr Ser Lys Ala Trp Lys Leu Asp Ala His Leu Cys Lys His
     50                  55                  60

Thr Gly Glu Arg Pro Phe Val Cys Asp Tyr Glu Gly Cys Gly Lys Ala
 65                  70                  75                  80

Phe Ile Arg Asp Tyr His Leu Ser Arg His Ile Leu Thr His Thr Gly
                 85                  90                  95

Glu Lys Pro Phe Val Cys Ala Ala Thr Gly Cys Asp Gln Lys Phe Asn
            100                 105                 110

Thr Lys Ser Asn Leu Lys Lys His Phe Glu Arg Lys His Glu Asn Gln
        115                 120                 125

Gln Lys Gln Tyr Ile Cys Ser Phe Glu Asp Cys Lys Lys Thr Phe Lys
    130                 135                 140

Lys His Gln Gln Leu Lys Ile His Gln Cys Gln His Thr Asn Glu Pro
145                 150                 155                 160

Leu Phe Lys Cys Thr Gln Glu Gly Cys Gly Lys His Phe Ala Ser Pro
                165                 170                 175

Ser Lys Leu Lys Arg His Ala Lys Ala His Glu Gly Tyr Val Cys Gln
            180                 185                 190

Lys Gly Cys Ser Phe Val Ala Lys Thr Trp Thr Glu Leu Leu Lys His
        195                 200                 205

Val Arg Glu Thr His Lys Glu Glu Ile Leu Cys Glu Val Cys Arg Lys
    210                 215                 220

Thr Phe Lys Arg Lys Asp Tyr Leu Lys Gln His Met Lys Thr His Ala
225                 230                 235                 240

Pro Glu Arg Asp Val Cys Arg Cys Pro Arg Glu Gly Cys Gly Arg Thr
                245                 250                 255

Tyr Thr Thr Val Phe Asn Leu Gln Ser His Ile Leu Ser Phe His Glu
            260                 265                 270

Glu Ser Arg Pro Phe Val Cys Glu His Ala Gly Cys Gly Lys Thr Phe
        275                 280                 285

Ala Met Lys Gln Ser Leu Thr Arg His Ala Val Val His Asp Pro Asp
    290                 295                 300

Lys Lys Lys Met Lys Leu Lys Val Lys Lys Ser Arg Glu Lys Arg Ser
305                 310                 315                 320

Leu Ala Ser His Leu Ser Gly Tyr Ile Pro Pro Lys Arg Lys Gln Gly
```

325 330 335

Gln Gly Leu Ser Leu Cys Gln Asn Gly Glu Ser Pro Asn Cys Val Glu
340 345 350

Asp Lys Met Leu Ser Thr Val Ala Val Leu Thr Leu Gly
355 360 365

<210> 3
<211> 1273
<212> DNA
<213> Human

<400> 3
atgcgcagca gcggcgccga cgcggggcgg tgcctggtga ccgcgcgcgc tcccggaagt 60

gtgccggcgt cgcgcgaagg ttcagcaggg agccgtgggc cgggcgcgcc ggttcccggc 120

acgtgtctcg gcacgtggca gcgcgcctgg ccctgggctt ggaggcgccg gcgccctgga 180

tccgccggcc gtggtcgccg agtcggtgtc gtccttgacc atcgccgacg cgttcattgc 240

agccggcgag agctcagctc cgaccccgcc gcgccccgcg cttcccagga ggttcatctg 300

ctccttccct gactgcagcg ccaattacag caaagcctgg aagcttgacg cgcacctgtg 360

caagcacacg ggggagagac catttgtttg tgactatgaa gggtgtggca aggccttcat 420

cagggactac catctgagcc gccacattct gactcacaca ggagaaaagc cgtttgtttg 480

tgcagccact ggctgtgatc aaaaattcaa cacaaaatca aacttgaaga aacattttga 540

acgcaaacat gaaaatcaac aaaaacaata tatatgcagt tttgaagact gtaagaagac 600

ctttaagaaa catcagcagc tgaaaatcca tcagtgccag cataccaatg aacctctatt 660

caagtgtacc caggaaggat gtgggaaaca ctttgcatca cccagcaagc tgaaacgaca 720

tgccaaggcc cacgagggct atgtatgtca aaaaggatgt tcctttgtgg caaaaacatg 780

gacggaactt ctgaaacatg tgagagaaac ccataaagag gaaatactat gtgaagtatg 840

ccggaaaaca tttaaacgca aagattacct taagcaacac atgaaaactc atgccccaga 900

aagggatgta tgtcgctgtc caagagaagg ctgtggaaga acctatacta ctgtgtttaa 960

tctccaaagc catatcctct ccttccatga ggaaagccgc ccttttgtgt gtgaacatgc 1020

tggctgtggc aaaacatttg caatgaaaca aagtctcact aggcatgctg ttgtacatga 1080

tcctgacaag aagaaaatga agctcaaagt caaaaaatct cgtgaaaaac ggagtttggc 1140

ctctcatctc agtggatata tccctcccaa aaggaaacaa gggcaaggct tatctttgtg 1200

tcaaaacgga gagtcaccca actgtgtgga agacaagatg ctctcgacag ttgcagtact 1260

taccttggc taa 1273

<210> 4
<211> 1213
<212> DNA
<213> Human

<400> 4

gtgccggcgc cgcgcgaagg ttcagcaggg agccgtgggc cgggcgcgcc ggttcccggc 60

acgtgtctcg gcacgtggca gcgcgcctgg ccctgggctt ggaggcgccg gcgccctgga 120

tccgccggcc gtggtcgccg agtcggtgtc gtccttgacc atcgccgacg cgttcattgc 180

agccggcgag agctcagctc cgaccccgcc gcgccccgcg cttcccagga ggttcatctg 240

ctccttccct gactgcagcg ccaattacag caaagcctgg aagcttgacg cgcacctgtg 300

caagcacacg ggggagagac catttgtttg tgactatgaa gggtgtggca aggccttcat 360

cagggactac catctgagcc gccacattct gactcacaca ggagaaaagc cgtttgtttg 420

tgcagccact ggctgtgatc aaaaattcaa cacaaaatca aacttgaaga aacattttga 480

acgcaaacat gaaaatcaac aaaaacaata tatatgcagt tttgaagact gtaagaagac 540

ctttaagaaa catcagcagc tgaaaatcca tcagtgccag cataccaatg aacctctatt 600

caagtgtacc caggaaggat gtgggaaaca ctttgcatca cccagcaagc tgaaacgaca 660

tgccaaggcc cacgagggct atgtatgtca aaaaggatgt tcctttgtgg caaaaacatg 720

gacggaactt ctgaaacatg tgagagaaac ccataaagag gaaatactat gtgaagtatg 780

ccggaaaaca tttaaacgca aagattacct taagcaacac atgaaaactc atgccccaga 840

aagggatgta tgtcgctgtc caagagaagg ctgtggaaga acctatacta ctgtgtttaa 900

tctccaaagc catatcctct ccttccatga ggaaagccgc ccttttgtgt gtgaacatgc 960

tggctgtggc aaaacatttg caatgaaaca aagtctcact aggcatgctg ttgtacatga 1020

tcctgacaag aagaaaatga agctcaaagt caaaaaatct cgtgaaaaac ggagtttggc 1080

ctctcatctc agtggatata tccctcccaa aaggaaacaa gggcaaggct tatctttgtg 1140

tcaaaacgga gagtcaccca actgtgtgga agacaagatg ctctcgacag ttgcagtact 1200

taccettggc taa 1213

<210> 5
<211> 34
<212> DNA
<213> Human

<400> 5
cggggtacca aaaatgcgca gcagcggcgc cgac 34

<210> 6
<211> 21
<212> DNA
<213> Human

<400> 6
tccttccctg actgcagcgc c 21

<210> 7
<211> 20
<212> DNA
<213> Human

<400> 7
tgcacaggtg cgcgtcaagc 20

<210> 8
<211> 20
<212> DNA
<213> Human

<400> 8
cacaaacaaa tggtctctcc 20

<210> 9
<211> 30
<212> DNA
<213> Human

<400> 9
cggtctagat tagccaaggg taagtactgc 30

<210> 10
<211> 30
<212> DNA
<213> Human

<400> 10
cctcccgggg ccaagggtaa gtactgcaac 30

Please type a plus sign (+) inside this box → [+]

PTO/SB/01 (8-96)
Approved for use through 9/30/98 OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

DECLARATION FOR UTILITY OR DESIGN PATENT APPLICATION

[X] Declaration Submitted with Initial Filing OR [ ] Declaration Submitted after Initial Filing

Attorney Docket Number	146.1364
First Named Inventor	BORDON-PALLIER et al
COMPLETE IF KNOWN	
Application Number	PCT/FR99/02738
Filing Date	11/9/99
Group Art Unit	
Examiner Name	

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled :

HUMAN htFIII GENE AND CODED htfIIIA PROTEIN

(Title of the Invention)

the specification of which

[ ] is attached hereto

OR

[X] was filed on (MM/DD/YYYY) Nov. 9, 1999 as United States Application Number or PCT International

Application Number PCT/FR99/02738 and was amended on (MM/DD/YYYY) _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment specifically referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37 Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35 United States Code §119 (a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate, or §365 (a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or of any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Number(s)	Country	Foreign Filing Date (MM/DD/YYYY)	Priority Not Claimed	Certified Copy Attached? YES	Certified Copy Attached? NO
98/14146	France	11/10/98	[ ]	[ ]	[ ]
PCT/FR99/02738	France	11/9/99	[ ]	[ ]	[ ]
			[ ]	[ ]	[ ]
			[ ]	[ ]	[ ]
			[ ]	[ ]	[ ]
			[ ]	[ ]	[ ]

[ ] Additional foreign application numbers are listed on a supplemental priority sheet attached hereto:

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below.

Application Number(s)	Filing Date (MM/DD/YYYY)

[ ] Additional provisional application numbers are listed on a supplemental priority sheet attached hereto.

146.1364
Please type a plus sign (+) inside this box → +

PTO/SB/01 (8-00)
Approved for use through 9/30/98. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

DECLARATION

I hereby claim the benefit under Title 35, United States Code §120 of any United States application(s), or §365(c) of any PCT international application designating the United States of America, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of Title 35, United States Code §112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application.

U.S. Parent Application Number	PCT Parent Number	Parent Filing Date (MM/DD/YYYY)	Parent Patent Number *(if applicable)*

☐ Additional U.S. or PCT international application numbers are listed on a supplemental priority sheet attached hereto.

As a named inventor, I hereby appoint the following registered practitioner(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Name	Registration Number	Name	Registration Number
Charles A. Muserlian	19,683		
Jordan B. Bierman	18,629		
Donald C. Lucas	31,275		
Bierman, Muserlian and Lucas	18,818		

☐ Additional registered practitioner(s) named on a supplemental sheet attached hereto.

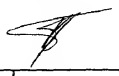
Direct all correspondence to:

Name	Bierman, Muserlian and Lucas				
Address					
Address	600 Third Avenue				
City	New York	State	New York	ZIP	10016
Country	U.S.A.	Telephone	(212) 661-8000	Fax	(212) 661-8002

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name of Sole or First Inventor: ☐ A petition has been filed for this unsigned inventor

Given Name	Middle Initial	Family Name	Suffix e.g. Jr.
FLORENCE		BORDON-PALLIER	

Inventor's Signature: X  Date: X 23 April 2001

Residence: City	State	Country	Citizenship
Guyancourt		France FRX	FR

Post Office Address	
Post Office Address	37, boulevard Beethoven

City	State	Zip	Country
Guyancourt		F-78280	France

☒ Additional inventors are being named on supplemental sheet(s) attached hereto

Please type a plus sign (+) inside this box → [+] 146.1364

PTO/SB/01 (8-96)
Approved for use through 9/30/98. OMB 0651-0032
Patent and Trademark Office U.S. DEPARTMENT OF COMMERCE

DECLARATION

ADDITIONAL INVENTOR(S) Supplemental Sheet

Name of Additional Joint Inventor, if any: ☐ A petition has been filed for this unsigned inventor

Given Name	CORINNE	Middle Initial		Family Name	ROCHER	Suffix e.g. Jr.	
Inventor's Signature	x [signature]					Date	x 23, April 2001
Residence: City	Paris	State		Country	France FRX	Citizenship	FR
Post Office Address							
Post Office Address	3, rue Elisa Lemonnier						
City	Paris	State		Zip	F-75012	Country	France

Name of Additional Joint Inventor, if any: ☐ A petition has been filed for this unsigned inventor

Given Name		Middle Initial		Family Name		Suffix e.g. Jr.	
Inventor's Signature						Date	
Residence: City		State		Country		Citizenship	
Post Office Address							
Post Office Address							
City		State		Zip		Country	

Name of Additional Joint Inventor, if any: ☐ A petition has been filed for this unsigned inventor

Given Name		Middle Initial		Family Name		Suffix e.g. Jr.	
Inventor's Signature						Date	
Residence: City		State		Country		Citizenship	
Post Office Address							
Post Office Address							
City		State		Zip		Country	

Name of Additional Joint Inventor, if any: ☐ A petition has been filed for this unsigned inventor

Given Name		Middle Initial		Family Name		Suffix e.g. Jr.	
Inventor's Signature						Date	
Residence: City		State		Country		Citizenship	
Post Office Address							
Post Office Address							
City		State		Zip		Country	

☐ Additional inventors are being named on supplemental sheet(s) attached hereto

(January 1997)